Towards vocal-behaviour and vocal-health assessment using distributions of acoustic parameters

ScuDo

Scuola di Dottorato ~ Doctoral School

WHAT YOU ARE, TAKES YOU FAR

Doctoral Dissertation

Doctoral Program in Metrology (30^{th}cycle)

Towards vocal-behaviour and vocal-health assessment using distributions of acoustic parameters

By

Antonella Castellana

******

Supervisor(s):

Prof. Alessio Carullo, Supervisor

Prof. Arianna Astolfi, Co-Supervisor

Doctoral Examination Committee:

Prof. Francesco Martellotta, *Referee*, Politecnico di Bari

Prof. Sten Ternström, *Referee*, KTH Royal Institute of Technology

Prof. Svante Granqvist, KTH Royal Institute of Technology and Karolinska Institutet

Prof. Daryush D. Mehta, Harvard Medical School, Massachusetts General Hospital and MGH Institute of Health Professions

Prof. Maria Södersten, Karolinska Institutet

Politecnico di Torino

2018

Declaration

I hereby declare that, the contents and organization of this dissertation constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data.

Antonella Castellana
2018

* This dissertation is presented in partial fulfillment of the requirements for **Ph.D. degree** in the Graduate School of Politecnico di Torino (ScuDo).

I would like to dedicate this thesis to my family

Acknowledgements

And I would like to express my gratitude to all of you, who have shared with me these PhD years.

Thanks Arianna and Alessio for your continuous support in my research and for trusting me since my first day here at Politecnico di Torino.

My special gratitude to Massimo Spadola Bisetti, Jacopo Colombini and all the ENT team at Molinette Hospital in Torino. Two chapters would have not been in this thesis without your expertise in voice and your availability to perform videostroboscopy examination.

A big thank you is to be addressed to all the researchers I met in KTH Royal Institute of Technology and in Karolinska University Hospital in Stockholm.

Thanks prof. Sten Ternström for your extreme availability in hosting me within your research group, for your warm welcome every morning and for your guidance.

Thanks prof. Maria Södersten for your interest in my research, for the time you spent with me to manage and perform experiments, and for your inviting me to the concerts!

Thanks prof. Svante Granqvist for your generosity, for your help in using voice analysis programs, for your availability as participant to experiments and, especially, for discussing with me all the results in a constructive way.

Thanks Andreas Selamtzis for your "mediterranean" welcome and friendship that made me feel at home. Thanks for your kind cooperation and constructive co-authorship.

Thanks Annika Szabo Portela, Anna Lundblad and Giampiero Salvi for your enthusiasm and support.

My sincere gratitude to my research group. Thanks Louena, Giusi, Lady, Pasquale, Elena, Giulia, Andrea, Alessandro, Sonja and Cristina for your guidance, support and friendship. Thank you Simone for your help in MATLAB and for our "working" lunches.

Thanks PoliEtnico, the choir I joined three years ago. Thank you for each shared moment during rehearsals and concerts. Thank you for all the choristers who participated in experiments with enthusiasm. A special thanks to Umberto for inviting me to join the choir, for his support in MATLAB and during experiments and, especially, for his friendship.

Thanks to all my friends. Mari and Zaira, you support me in good and bad moments. Thanks Giangi for your power of make me smile in every situation. Thanks Simona, Anna, Michele and Wanda for your continuous presence in my life.

A big thank you to my family. Your patience, your calls and messages supported me every days. Thanks Carlo, you faced all the difficulties with me during these years and you helped me a lot.

My final thanks is to the light and column of my days. Thank you Gas for your love and for being always close to me.

Abstract

Voice disorders at different levels are affecting those professional categories that make use of voice in a sustained way and for prolonged periods of time, the so-called occupational voice users. In-field voice monitoring is needed to investigate voice behaviour and vocal health status during everyday activities and to highlight work-related risk factors. The overall aim of this thesis is to contribute to the identification of tools, procedures and requirements related to the voice acoustic analysis as objective measure to prevent voice disorders, but also to assess them and furnish proof of outcomes during voice therapy.

The first part of this thesis includes studies on vocal-load related parameters. Experiments were performed both in-field and in laboratory. A one-school year longitudinal study of teachers' voice use during working hours was performed in high school classrooms using a voice analyzer equipped with a contact sensor; further measurements took place in the semi-anechoic and reverberant rooms of the National Institute of Metrological Research (I.N.Ri.M.) in Torino (Italy) for investigating the effects of very low and excessive reverberation in speech intensity, using both microphones in air and contact sensors. Within this framework, the contributions of the sound pressure level (SPL) uncertainty estimation using different devices were also assessed with proper experiments. Teachers adjusted their voice significantly with noise and reverberation, both at the beginning and at the end of the school year. Moreover, teachers who worked in the worst acoustic conditions showed higher SPLs and a worse vocal health status at the end of the school year. The minimum value of speech SPL was found for teachers in classrooms with a reverberation time of about 0.8 s. Participants involved into the in-laboratory experiments significantly increased their speech intensity of about 2.0 dB in the semi-anechoic room compared with the reverberant room, when describing a map. Such results are related to the speech monitorings performed with the vocal analyzer, whose uncertainty estimation for SPL differences resulted of about 1 dB.

The second part of this thesis was addressed to vocal health and voice quality assessment using different speech materials and devices. Experiments were performed in clinics, in collaboration with the Department of Surgical Sciences of Università di Torino (Italy) and the Department of Clinical Science, Intervention and Technology of Karolinska Institutet in Stockholm (Sweden). Individual distributions of Cepstral Peak Prominence Smoothed (CPPS) from voluntary patients and control subjects were investigated in sustained vowels, reading, free speech and excerpted vowels from continuous speech, which were acquired with microphones in air and contact sensors. The main influence quantities of the estimated cepstral parameters were also identified, which are the fundamental frequency of the vocalization and the broadband noise superimposed to the signal. In addition, the reliability of CPPS estimation with respect to the frequency content of the vocal spectrum was evaluated, which is mainly dependent on the bandwidth of the measuring chain used to acquire the vocal signal. Regarding the speech materials acquired with the microphone in air, the 5th percentile resulted the best statistic for CPPS distributions that can discriminate healthy and unhealthy voices in sustained vowels, while the 95th percentile was the best in both reading and free speech tasks. The discrimination thresholds were 15 dB (95% Confidence Interval, CI, of 0.7 dB) and 18 dB (95% CI of 0.6 dB), respectively, where lower values indicate a high probability to have unhealthy voice. Preliminary outcomes on excerpted vowels from continuous speech stated that a CPPS mean value lower than 14 dB designates pathological voices. CPPS distributions were also effective as proof of outcomes after interventions, e.g. voice therapy and phonosurgery. Concerning the speech materials acquired with the electret contact sensor, a reasonable discrimination power was only obtained in the case of sustained vowel, where the standard deviation of CPPS distribution higher than 1.1 dB (95% CI of 0.2 dB) indicates a high probability to have unhealthy voice. Further results indicated that a reliable estimation of CPPS parameters is obtained provided that the frequency content of the spectrum is not lower than 5 kHz: such outcome provides a guideline on the bandwidth of the measuring chain used to acquire the vocal signal.

Contents

List of Figures

List of Tables

Nomenclature

Subscripts

ref reference microphone

Acronyms / Abbreviations

$\overline{s}$ the average of s_{i} values

CI Confidence Interval

CPP Cepstral Peak Prominence

CPPS Cepstral Peak Prominence Smoothed

$CPPS_{\mathrm{5prc}}$ fifth percentile of CPPS distribution

$CPPS_{\mathrm{95prc}}$ ninety-fifth percentile of CPPS distribution

$CPPS_{\mathrm{kurt}}$ kurtosis of CPPS distribution

$CPPS_{\mathrm{mean}}$ mean of CPPS distribution

$CPPS_{\mathrm{median}}$ median of CPPS distribution

$CPPS_{\mathrm{mode}}$ mode of CPPS distribution

$CPPS_{\mathrm{range}}$ interval of CPPS distribution

$CPPS_{\mathrm{skew}}$ skewness of CPPS distribution

$CPPS_{\mathrm{std}}$ standard deviation of CPPS distribution

CVPs Conversational Voice Parameters

$Dt\%$ phonation time percentage

ECM Electret Condenser Microphone

$F0$ fundamental frequency

$GRBAS$ Grade, Roughness, Breathiness, Asthenia and Strain

GUM Guide to the expression of Uncertainty in Measurement

$HATS$ Head and Torso Simulator

L_{A90} the A-weighted noise level that is exceeded by 90% of the sample

LME Linear Mixed Effects model

$LTAS$ Long Term Average Spectrum

MIC microphone in air

$OVPs$ Occupational Voice Parameters

PM piezoelectric contact microphone

rms root mean square

$s(g)$ the experimental standard deviation of each device-group

s_{i} the experimental standard deviation of repeated measures

s_{m} the standard deviation of the mean

SD standard deviation

SE standard error

SPL Sound Pressure Level

SPL_{eq} equivalent SPL

SPL_{mean} mean of SPL distribution

SPL_{mode} mode of SPL distribution

$T30_{0.25\div 2\mathrm{kHz,occ}}$ average reverberation time between 0.25 and 2 kHz

WHO World Health Organization

Chapter 1

Introduction

Dysphonia is an oral communication disorder of the voice that impedes an individual from expressing their verbal and emotional message [1]. As such, dysphonia may have an impact on the individual's quality of life, thus constituting an activity limitation and/or participation restriction. About one third of the labour force works in professions in which the voice is the primary tool [2] and it has been shown that individuals in high-voice use occupations are more likely to develop voice disorders than in other occupations [3, 4]. The association between the occurrence of voice symptoms and occupational voice use has been reported in the 58% of cases for teachers and in the 29% of cases for other occupational voice users [5].

1.1 Vocal loading

Many voice disorders are chronic or recurring conditions resulting from abusive patterns of vocal behaviour with, as a consequence, vocal fold tissue reactions to mechanical stress. *Vocal load* is defined as a combination of prolonged voice use and additional factors, such as elevated phonation frequency and high sound pressure level [6–8]. As an intensive physiological voluntary activity, intensively speaking requires a certain effort. *Vocal effort* is a physiological magnitude that accounts for changes in voice production induced by the distance from the listeners, noise and the physical environment [9]. Vocal load is described using the following three parameters, as suggested by its definition: voice Sound Pressure Level (SPL),

fundamental frequency (F0) and vocal dose. Different vocal doses types have been defined by Titze, Ŝvec, Popolo *et al.* [10, 11]:

- *time dose*, the total phonation time;
- *voicing time percentage*, the percentage of time spent phonating for the total monitoring period;
- *cycle dose*, the entire vocal folds vibration cycle;
- *distance dose*, the total distance travelled by vocal folds including F0, phonation time and SPL;
- *energy dissipation dose*, the amount of heat produces by vocal folds;
- *radiated energy dose*, all the energy emanated from the mouth.

1.1.1 Long-term voice monitoring

Voice disorders at different levels are affecting those professional categories that make use of voice in a sustained way and for prolonged periods of time (e.g. actors, singers, call-center employees, sales people) [12–14]. The appearance of voice disorders may bring to absenteeism from work in order to recover [15], with a consequent impact on the economy in terms of health care use, voice-related absence and productivity loss at work [16]. Differently from in-laboratory measurements, the use of specific tools and practices to in-field monitorings of voice, e.g., by means of vocal analyzers [17–19], combined with objective environmental measurements, can offer insight into the changes of vocal loading during working hours and can help to identify a person's risk of vocal dysfunctions [20, 21]. Moreover, portable vocal analyzers have recently allowed the relationships between daily vocal load and voice disorders to be investigated [22, 23]. Although many works using portable voice analyzers have been conducted, see Szabo Portela [24] for a summary, further long-term monitorings of voice are needed to characterize the vocal behaviour of occupational voice users during working activities.

1.1.2 Portable voice analyzers

In the last two decades, different portable voice analyzers, also called voice accumulators or voice dosimetry devices, have been developed. They are all equipped with a contact sensor that allows to minimize the effects of sound sources different from the voice of interest. The NCVS dosimeter, developed at the National Centre for Voice and Speech for research (Denver, CO, US) [25, 26], and the Ambulatory Phonation Monitor, APM, (KayPENTAX, Montvale, NJ, USA), developed by the Massachusetts General Hospital [27, 28], both use an accelerometer to sense the skin acceleration due to the vibration of the vocal folds. The VoxLog, a commercial device developed at the Linköpings University of Sweden (Sonvox AB, Umeä, Sweden) [29], is provided with a miniature accelerometer and a microphone that allow measuring both voice level and environmental noise. The Voice Care device (PR.O.VOICE, Turin, Italy), recently developed at Politecnico di Torino, uses an electret condenser microphone as a contact sensor [30].

Voice Care

The Voice Care device consists of a data-logger connected to an Electret Condenser Microphone (ECM AE38 [Alan Electronics GmbH (Dreieich, Germany)]), which is fixed at the jugular notch by means of a surgical band, thus sensing the skin vibrations induced by the vocal-fold activity (Figure 1.1). Figure 1.2 represents the block scheme of the Voice Care device. The output signal of the ECM is suitably conditioned through an analogue circuitry in order to match its characteristics (amplitude and frequency content) to the analogue-to-digital converter internal to a micro-controller based board. The samples acquired with the ECM are grouped into frames of 30 ms and only voiced frames are processed [31]. The choice of such interval arises from the evidence that the minimum duration of pauses in Italian readings is equal to 60 ms [32], but pause lengths of 30 ms can also occur in storytelling style speech [33], so that a 30 ms-interval guarantees an effective discrimination between voiced and unvoiced frames. This frame duration is also used in other dosimeters that are equipped with contact sensors [34, 25, 26]. The raw samples are stored on an internal memory device (SD card) and then post-processed with suitable software programmes on a PC.

Fig. 1.1 The Voice Care device.

The Voice Care measures the speech SPL of the speaker at a fixed distance (in dB), the F0 (in Hz) and the phonation time percentage (Dt% in %), which is defined as the percentage of time spent phonating for the total monitoring period [11]. The results related to F0 and SPL are usually shown as histograms of occurrences that allow revealing important characteristics of the vocal behaviour over many hours.

In order to estimate the speech SPL of the speaker at a fixed distance d_0 of 16 cm in front of the mouth, each subject has to perform a preliminary calibration,

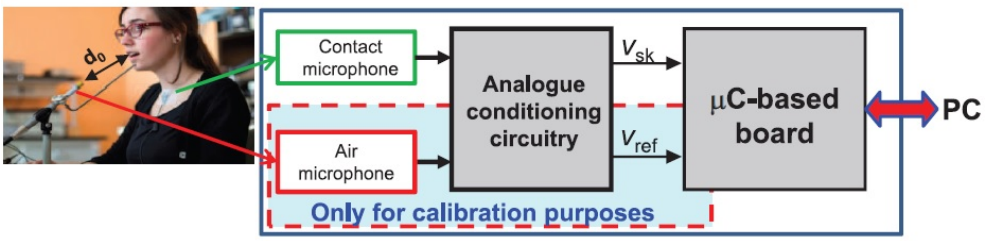

Fig. 1.2 Block scheme of the Voice Care device.

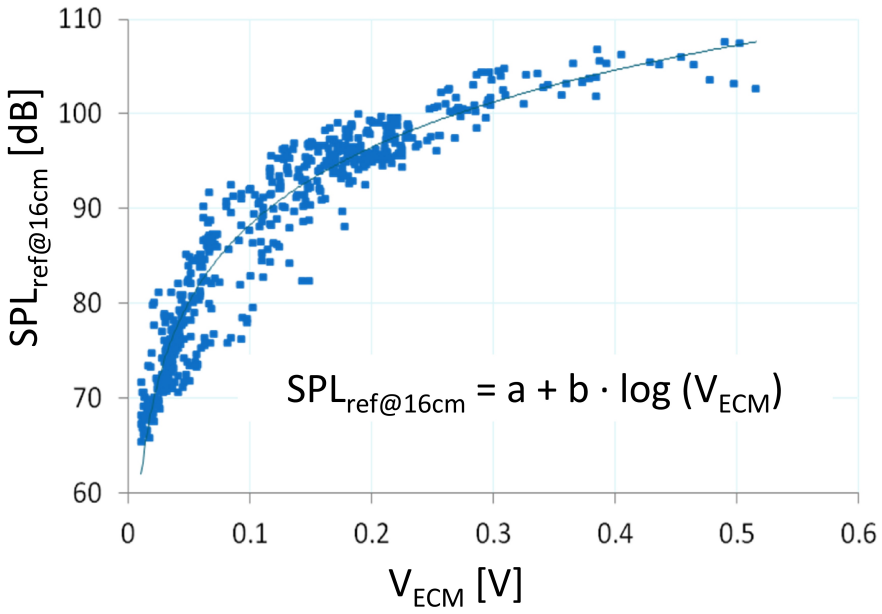

Fig. 1.3 The calibration function, where $SPL_{\text{ref@16cm}}$ is the sound pressure level measured at the microphone in air (dB) and the V_{ECM} is the voltage signal acquired at the contact microphone placed at the jugular notch (V).

repeating the vowel /a/ at increasing levels in front of a microphone in air (Behringer ECM8000), used as a reference (Figure 1.2). The samples of the ECM signal and the signal at the output of the reference-microphone chain are acquired, grouped into the fixed-length frames and then processed in order to estimate the root mean square (*rms*) values V_{ECM} and V_{ref} for each frame, respectively. Such procedure, which is needed before starting each monitoring, is designed to identify the function that relates the voltage signal at the output of the ECM chain, V_{ECM}, to the reference SPLs at the fixed distance from the mouth of the subject under monitoring, $SPL_{\text{ref@16cm}}$, as shown in Figure 1.3 [31]. For each monitoring using Voice Care, SPL values for voiced frames at a fixed distance of 16 cm from the speaker's mouth are then obtained, thanks to the calibration function estimated for each subject.

1.1.3 Investigations on teachers

Several authors have investigated the prevalence of voice problems in teachers [4, 35, 36] and all have agreed that vocal loading and vocal effort are two of the most important causes of voice dysfunctions [37, 2, 38]. Several studies have used

a portable voice analyzer to assess SPL, F0, and the phonation time: they were significantly higher in teaching situations compared with non-teaching situations [39, 40], thus highlighting a risky situation for teachers at work. Up to 50% of teachers reported having suffered adverse vocal symptoms at least once in their career [36].

In the existing literature, only few studies have been conducted on long-term voice monitoring during teachers' working activity and such studies mainly investigated the objective changes that took place in teachers' voice parameters over a period of one working day. Rantala *et al.* [41] investigated voice changes during a working day in 33 primary and secondary school teachers using voice recording methods. The teachers recorded the first and the last lesson during a working day with a digital audiotape recorder and a head-mounted microphone located on the side of their mouth. The most evident change during a working day was an increase of 7.9 Hz in the mean value of F0 from the first to the last lesson. Laukkanen *et al.* [38] evaluated the vocal fatigue of 47 primary school teachers over one working day using a portable digital recorder and a microphone attached to the headset. They found by comparing the vocal data recorded at the beginning and at the end of the working day that teachers showed a higher SPL and a higher F0 at the end of the working activity. Hunter and Titze [42] used data from the National Center for Voice and Speech (NCVS) vocal data bank, collected by means of the NCVS Voice Dosimeter [43] over a period of two working weeks, to study the voice use of 57 teachers during occupational activities. They found that the F0 appeared to trend upwards throughout the working day, and that the teachers experienced a wide range of voicing time percentages (Dt%) (30% $\pm$ 11%). Increased F0 and SPL, following vocal loading, have been interpreted in different studies as an adaptation to vocal loading [41, 38, 43, 44].

Changes in voice production can be induced by environmental factors, such as the noise level. Södersten *et al.* [45] used a DAT recorder to measure voice SPL, F0, and phonation time on ten female preschool teachers during teaching situations and in a separately performed reading task without background noise: a 9 dB louder SPL and a higher mean F0 in the teaching situation compared with the reading task were found. The involuntary tendency of speakers to increase their voice level as the noise level increases in order to improve intelligibility of the speech signal is called *Lombard effect* [46–49]. Lane and Tranel [46] summarized a wide range of findings reported in the literature about this effect. Lazarus [47] found that the

speech level rises as the noise level rises with a slope of 0.3–0.6 dB for each 1 dB of increase in the mean value of the A-weighted noise level distribution above 45 dB. Bottalico and Astolfi [48] and Sato and Bradley [49] studied the vocal parameters of primary school teachers in relation to activity noise levels; they found a growing rate of the speech level with the noise level of 0.7 dB/dB. To the best of the author knowledge, the relationship between the activity noise levels and voice parameters of secondary school teachers has not yet been studied. As far as the activity noise levels in secondary schools are concerned, only Shield *et al.* [50] measured the noise levels during teaching activities. They found L_{A90} background noise levels (i.e., the A-weighted noise level that is exceeded by 90% of the sample) to be between 38 and 63 dB, with a mean value of 51 dB (standard deviation=6 dB). However, Shield *et al.* [50] and other authors [51–54], who reported the activity noise level in primary schools, only documented the activity noise conditions over one period during the school year. Therefore, it is not known whether the noise conditions remain unchanged during the course of the year, or whether prolonged exposition to high noise levels has any effect on the students' behaviour, such as whether students make more noise or less noise.

Another important factor that should be taken into account to evaluate voice production under realistic communication conditions is the effect of room acoustics. Brunskog *et al.* [55] and Pelegrín-García *et al.* [56] found that the average voice level of speakers is closely related to the "room gain," which represents the gain that is given to the speaker's voice due to the reflections in the room. Room gain has been found to be closely correlated to reverberation time [57]. Brunskog *et al.* [58] found a variation in the voice power level at a rate of -13.5 dB per 1 dB of increase in room gain. A tendency to lower the voice level when the room gain increased was also found in a study conducted by Pelegrín-García *et al.* [56]. They investigated the vocal effort of 13 male subjects under four different acoustic room conditions, with a reverberation time that ranged from 0.04 to 5.38 s. They found that talkers tend to vary their voice power level at a rate of -3.6 dB per 1 dB of room gain. Bottalico and Astolfi [48] monitored the voice parameters of 41 primary school teachers over one working week, and found that SPL mean @ 1 m and reverberation time are related by a quadratic regression curve, which shows a minimum value in correspondence to an average mid-frequency reverberation time of 0.8 s in occupied conditions.

The above mentioned studies have shown that the teachers' voice level depends on both noise and reverberation. Although noise and reverberation are simultaneously

present during teaching activities, studies that evaluate the simultaneous presence of both parameters and their combined effect on teachers' voice use have not yet been carried out. Moreover, all the studies on teachers' voice parameters have been conducted over a short period of time, i.e., no more than two weeks of working activities, or changes of teachers' vocal use has been evaluated during the course of only one working day. There is a lack of longitudinal studies with repeated measures to assess changes in teachers' vocal behaviour or noise conditions during teaching hours over a long period, such as one entire school year.

1.1.4 Uncertainty issues

Despite the large use of the above-mentioned vocal-load related parameters, researches do not usually take into account the uncertainty of measures when they report and discuss the results. In the existing literature, only few studies deal with the uncertainty estimation of speech SPL. Measurements of speech SPL with microphones in air mounted in front of the speaker bring to uncertain results due to the possible variation in the subject-to-microphone distance during the speech performance [59], while uncertain results for both headworn microphones and contact-sensor based devices are mainly due to calibration issues and more generally to a not well defined metrological characterization [31]. Moreover, speech SPL measures obtained from all the devices are also affected by uncertainty due to the variability of the speech itself [60]. A first attempt in estimating the average speech SPL measurement error using a contact-sensor based device was made by Hillman *et al.* [28], who compared SPL measures estimated with an accelerometer to SPL values provided by a reference microphone in air. Maximum SPL average error of 3.2 dB (standard deviation 6.0 dB) was obtained during different speech tasks after a proper calibration of the device was performed. An in depth study has been carried out by Švec *et al.* [18], who estimated the uncertainty of speech SPL values obtained by an accelerometer-based device, taking into account contributions related to the calibration function and to speech variability. They found that 30-ms frame speech SPL (i.e. instantaneous) at 30 cm from the speaker's mouth can be estimated from the skin acceleration level with a 95% confidence interval of ± 5 dB for female and ± 6 dB for males. Mean and equivalent speech SPL can instead be estimated with an accuracy better than 4.3 dB and 2.5 dB, respectively, in 95% of the cases of "normal" to "loud" speech [61]. Carullo *et al.* [31] took into account

calibration and instrumental uncertainty, repeatability and reproducibility for the Voice Care device, thus estimating a standard uncertainty for the instantaneous SPL at 16 cm from the speaker's mouth not greater than 2.3 dB, for male speakers, and 4.2 dB, for female speakers. In addition, the measurement error of instantaneous SPL during continuous speech was estimated against a reference microphone: a mean error of -0.8 dB and a standard deviation of 2.2 dB were obtained, thus confirming the estimated uncertainty. Even though the above-mentioned studies allowed the contact-sensor based devices to be preliminarily characterized from a metrological point of view, the variability contributions were obtained through experiments that involved human subjects, thus not allowing the effect of the reproducibility of the speech itself, i.e. the source reproducibility, to be distinguished from other possible causes of uncertainty.

1.2 Vocal health

Dysphonia appears when the vocal folds vibrate abnormally [62] and it involves abnormal *voice quality* much more frequently than abnormal pitch or loudness [63]. In order to investigate the health status of the vocal apparatus, the voice quality assessment is then needed. Several definitions of voice quality have been expressed as variation of the overall quality (or timbre) of a sound, which has been defined by the American National Standards Institute [64] as "that attribute of auditory sensation in terms of which a listener can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar" [65]. As such, voice quality is multidimensional.

The following list summarizes the protocol that is usually followed to explore the multifactorial voice nature during a clinical visit [66].

- *Careful case history*. This phase mainly consists in a patient's self reporting, as part of the nonstandardized clinical interview process. As such, it is subjective and likely to be unreliable, since commonly a daily voice use and misuse becomes routine and therefore it is carried out without awareness [67]. An important improvement have been obtained with the use of standardized self-report inventories such as the Voice-Related Quality of Life [68], the Voice

Handicap Index [69] and the Voice Activity and Participation Profile, which also addresses activity limitation and participation restriction [70].

- *Auditory perceptual assessment*. Since voice quality is auditory-perceptual by nature, its primary measurement technique is the auditory-perceptual rating scale. Several voice quality rating protocols have been introduced, such as the GRBAS interval scales (the acronym for Grade, Roughness, Breathiness, Asthenia and Strain) [71], the Stockholm Voice Evaluation Approach (SVEA) visual analog scales [72] and the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V) hybrid visual analog scales [73]. Different voice qualities are considered into such auditory-perceptual rating scales: for example, breathiness is present when vocal folds do not come fully together and the space between them allows air to flow through, thus provoking noise and turbolence; creakyness results from vocal folds very shortened and slackened that follow a very complex vibration patter and sound like a creaking door; strain comes from a high degree of muscle tension in the entire vocal tract that makes voice effortful [74]. The combination of breathiness and roughness is described in terms of hoarsness [75].

- *Videostroboscopy examination*. Videoendoscopy with stroboscopy (videostroboscopy) is the most important clinical tool for instrumental voice assessment: it allows for a direct observation of vocal fold anatomy and physiology thus constituting a precious method for identify voice disorders. [76]. However, such examination is intrusive, instantaneous and can only be performed in clinics, where patients have a different vocal behaviour from daily life.

- *In-clinic recordings of voice*. Since voice quality is an attribute of the output signal of the (normal or abnormal) vocal fold oscillations, numerous measuring techniques have been proposed. A summary of the acoustic analysis used for voice quality assessment will be outlined in the next paragraph. However, a common limitation affects in-clinic recordings of voice, i.e. they are performed in a totally different context from the work-place.

As highlighted by Manfredi *et al.* [67], clinicians need to know how their patients' voices sound in everyday life in order to identify defective patterns and try to modify them. The accent is no longer on vocal load (paragraph 1.1), but on voice

quality, which shows behavioural changes that are considered of great influence in the pathogenesis of many common voice disorders [77].

1.2.1 Acoustic parameters

Many algorithms and methods to obtain an objective analysis of dysphonia and its severity have been implemented (see Buder for an overview [78]).

The first investigated parameters were those in the time domain, e.g. jitter and shimmer, whose main limitations have been highlighted in the existing literature. Since they depend on the accurate identification of cycle boundaries, that is where a cycle of vocal-fold vibration starts and finishes, they become unreliable with highly perturbed signals [79]. Furthermore, the good performance of the speech task, i.e. a vowel produced with steady pitch and loudness, is very important for the computation of such parameters, since any changes in the signal could be read as increases in vocal perturbation [80].

To overcome such limitations, spectral- and cepstral-based measures are currently considered: they can be applied also to continuous speech that is able to represent everyday speaking patterns [81]. In particular, cepstral parameters have been defined as the most promising indexes of dysphonia severity. They are evaluated in the cepstrum domain, that is a log power spectrum of a log power spectrum [82]: while the first power spectrum shows the frequency distribution of the signal energy, the second spectrum indicates how periodic the harmonic components in the spectrum are.

Two cepstral parameters have been defined, namely the Cepstral Peak Prominence (CPP) and its smoothed version (CPPS). CPP is a measure (in dB) of the cepstral peak amplitude, normalized for overall signal amplitude through a linear regression line estimated relating quefrency to cepstral magnitude [83]. CPPS considers two smoothing steps before calculating the cepstral peak prominence [82]. The meta-analysis on correlation coefficients between acoustic measurements and perceptual evaluation of voice quality by Maryn *et al.* [84] highlighted the relevance of CPPS: they found that CPPS satisfied the meta-analytic criteria in sustained vowels as well as in continuous speech. CPPS has also resulted well correlated with perceptual judgement of overall grade of dysphonia and different types of voice quality [85, 86].

Additionally, significantly different CPPS values between dysphonic and control group have been found in the vowel /a/ [87]. Despite the attention given to the parameter, in the existing literature there is a lack of investigation on CPPS diagnostic precision. Such analysis has been performed for the Acoustic Voice Quality Index (AVQI), which is a multivariate construct that includes CPPS and other four acoustic metrics [88]. All the above-mentioned studies used cepstrum software packages to estimate CPPS, which only provide the mean of CPPS values and in some cases the standard deviation: the most popular packages are Praat [89], SpeechTool [90] and the Analysis of Dysphonia in Speech and Voice module [91] of Multi-Speech from KayPENTAX (Montvale, NJ). These programs process signals acquired with microphones in air only.

A recent work by Mehta *et al.* [92] evaluated CPP from vowels acquired with a microphone in air and an accelerometer sensor using a commercially available program. They found that CPP measures from the two sensors were highly correlated, without significant differences between healthy and pathological voice.

1.2.2 Wearable voice monitoring systems using smartphones

The vocal-behaviour assessment of occupational voice users was a tricky task when devices able to collect vocal parameters during long-term monitorings were unavailable. A first attempt to overcome this problem was made with the development of the wearable vocal analysers equipped with contact sensors that have been already described in 1.3, but these devices only provide vocal-load related parameters and their high cost prevents from monitoring a large number of subjects. The Portable Voice Lab [93] differs from the above-mentioned devices as it is equipped with a small microphone and provides not only measures of F0, but also voice quality indexes related to F0 irregularities and hoarseness. A recently proposed solution consists in using smartphone devices to collect voice data. The main advantage of these devices is that they offer an easy and low-cost way for performing repeated measurements over time, which is essential for baseline designs and for voice monitoring. Manfredi *et al.* [94] tested the reliability of commercial smartphones in assessing voice quality using synthesized voice samples with three levels of jitter and three levels of added noise. Mehta *et al.* investigated vocal-load measures [95] and also glottal airflow parameters [19] in signals acquired using a neck-placed miniature accelerometer as voice sensor and a custom smartphone application (Voice Health Monitor, VHM) as

data acquisition platform. They also made a comparison between some voice quality indexes, namely time-domain perturbation parameters and cepstral measures, computed on signals from the neck-placed miniature accelerometer and a microphone [92]. A mobile application for iOS devices [96] able to measure F0, jitter and CPP was also used in patients practice and the participants found numeric CPP feedback helpful in self-evaluating voice quality.

The diffusion of long-term monitorings instead of in-clinic short-term measurements has been providing distributional parameters that, differently from average measures [97], are able to detect patients with aberrant vocal behaviors that are related to voice disorders [98]. Based on these premises, Vocal Holter App (PR.O.VOICE, Turin, Italy), a smartphone application combined with a cheap contact microphone embedded in a collar, allows short and long-term monitoring and provides most of the results in terms of distributions of vocal parameters related to vocal load and voice quality [99].

1.3 Thesis at a glance: motivations and aims

The WHO defines activity limitations as difficulties an individual has in performing tasks, and participation restrictions as difficulties a person has in facing life situations [100]. Dysphonia is an oral communication disorder of the voice that impedes an individual from expressing their verbal and emotional message [1]; as such, dysphonia may have an impact on the individual's quality of life. About one third of the labour force works in professions in which the voice is their primary tool [2] and it has been shown that individuals in high-voice use occupations are more likely to develop voice disorders than in other occupations [3], [4]. Several researches have investigated the prevalence of voice problems in teachers [4, 35, 36] and all have agreed that vocal loading and vocal effort are two of the most important causes of voice dysfunctions. Recently, portable voice dosimeters equipped with contact microphones or accelerometers that sense the vocal fold vibrations have been developed to measure speech variations in terms of intensity, fundamental frequency, and phonation time, the so-called vocal loading. The importance of such devices is related to the possibility to collect voice data during working activities. However, few studies have been conducted on long-term voice monitoring during working

hours and the longest period of investigation was two weeks [39]. Therefore, the first objective of this thesis is

1. **To investigate teachers' vocal behaviour and to study the relationships between voice use and classroom acoustic parameters, through in-field longitudinal observations over a school year.**

Chapter 2 mainly describes the methodology and includes the obtained results as answers to this research aim.

Among the categories of risk factors for occupational voice problems, environmental factors have a relevant role. Numerous studies have dealt so far with changing in speech production for talkers due to different acoustic environments, but they have often been focused on the effect of noise or distance from the listeners [101]. Few data have been published reporting details on speech modifications while speaking in the presence of reverberation and such investigations used microphone in air only to acquire voice signals [102, 103]. The use of portable voice dosimeters in in-laboratory studies is needed in order to assess their outputs in controlled situations and better understand the measures from in-field monitorings. Moreover, there is a lack of information on the uncertainty related to vocal parameters estimated with these devices. Therefore, the second objective of this thesis is

2. **To investigate differences in speech intensity in very low and very high reverberant rooms, accounting for the uncertainty of the parameters estimated using a headworn microphone and a vocal analyzer.**

Chapter 3 presents results on the variability of sound pressure level estimated using three devices, when subjects speak at a comfortable level. These findings are essential for the investigation described in Chapter 4, which includes the obtained results as answers to this research aim.

There is the evidence that voice behaviour has the principal role in the pathogenesis of many common voice disorders [77], then clinicians are interested in how patients daily use their voices in order to identify defective patterns and try to modify them. Such an investigation differs from voice dosimetry, as it mainly contributes

with a-posteriori information about the voice use [67], and it is related to parameters on voice quality. In summary, there is the need of supporting short-term in-clinic recordings with long-term in-field monitorings that are able to characterize vocal health and vocal behaviour of patients during everyday activities. Among several acoustic markers proposed to objectify dysphonia type and severity from signals acquired with microphones in air, the Cepstral Peak Prominence Smoothed (CPPS) has been labeled as the most pertinent [84]. Therefore, the third objective of this thesis is

3. **To validate CPPS distributions as vocal health indicator in sustained vowels and continuous speech using microphones in air and contact sensors.**

Chapter 5 and chapter 6 include all the investigations performed on CPPS distributions: the diagnostic precision of descriptive statistics from CPPS distributions obtained from the different speech materials and devices, the best threshold values between healthy and pathological talkers and the respective variabilities, the repeatability of such measures and the association with perceptual ratings are the main evaluated aspects. Chapter 7 shows a comparison between cepstral and entropy analyses in excerpted vowels from readings of healthy and pathological voices.

Eventually, Chapter 8 summarizes the main findings of this thesis and includes recommendations for future research.

1.4 List of publications

As a result of the research that has been carried out within these three years of Ph.D., the following publications on international scientific journals were produced:

1. A. Castellana, A. Carullo, A. Astolfi, G. E. Puglisi, and U. Fugiglando, *Intra-speaker and inter-speaker variability in speech sound pressure level across repeated readings*, J. Acoust. Soc. Am., 141(4), 2353–2363 (2017).

2. G. Calosso, G.E. Puglisi, A. Astolfi, A. Castellana, A. Carullo, F. Pellerey, *A 1-school year longitudinal study of secondary school teachers' voice parameters*

and influence of classroom acoustics, J. Acoust. Soc. Am., 142 (2), 1055-66 (2017).

3. A. Castellana, A. Carullo, S. Corbellini and A. Astolfi, *Discriminating pathological voice from healthy voice using Cepstral Peak Prominence Smoothed distribution in sustained vowel*, IEEE Transactions on Instrumentation and Measurement, accepted.

4. A. Astolfi, A. Castellana, A. Carullo, and G. E. Puglisi, *Measurement uncertainty of speech level and speech level difference for a contact-sensor-based device and a headworn microphone*, J. Acoust. Soc. Am., submitted.

5. A. Astolfi, A. Castellana, G. E. Puglisi, U. Fugiglando, and A. Carullo, and *Investigation on the effects of very low and excessive reverberation in speech levels*, J. Acoust. Soc. Am., to be submitted.

Others works allowed discussing first results of the research and to look forward with further steps:

- G.E. Puglisi, L.C. Cantor Cutiva, L. Pavese, A. Castellana, M. Bona, S. Fasolis, V. Lorenzatti, A. Carullo, A. Burdorf, F. Bronuzzi, A. Astolfi, *Acoustic comfort in highschool classrooms for students and teachers*, Energy Procedia 78, 3096-3101 (2015)

- A. Castellana, A. Carullo, F. Casassa, A. Astolfi, L. Pavese, G. E. Puglisi, *Performance comparison of different contact microphones used for voice monitoring*, 22th International Congress on Sound and Vibration – ICSV 2015, Firenze (Italy), 12-16 July 2015

- A. Castellana A., F. Casassa, G.E. Puglisi,*Nuovi parametri acustici utuli nella diagnostica e nella prevenzione di patologie vocali*, 42° Convegno Nazionale AIA, Firenze (Italy), 16-17 July 2015

- F. Casassa, A. Castellana, G.E. Puglisi, *Confronto tra sensori a contatto per il monitoraggio vocale*, 42° Convegno Nazionale AIA, Firenze (Italy), 16-17 July 2015

- A. Astolfi, A. Carullo, A. Castellana, G. Calosso, G.E. Puglisi, M. Spadola Bisetti, A. Accornero, S. Corbellini, P. Bottalico, L. Pavese, R. Albera, *Voice*

Care®: un "Holter vocale" per il monitoraggio dei professionisti della voce, 22° Convegno di Igiene Industriale "Le Giornate di Corvara" (Italy), 30 March-1 April 2016

- A. Astolfi, G.E. Puglisi, A. Carullo, A. Castellana, U. Fugiglando, *Effetti di scarsa ed eccessiva riverberazione sul parlato continuo*, 43° Convegno Nazionale Associazione Acustica Italiana – AIA 2016, Alghero (Italy), 25-27 May 2016

- G. Calosso, G. E. Puglisi, A. Astolfi, A. Castellana, A. Carullo, F. Pellerey, *Relationships between classroom acoustics and voice parameters of teachers at the beginning and at the end of a school year*, EuroRegio2016 – Porto (Portugal), 13-15 June 2016

- A. Astolfi, G. E. Puglisi, A. Castellana, A. Carullo, U. Fugiglando, *Speech level in rooms with very low and very high reverberation*, EuroRegio2016 – Porto (Portugal), 13-15 June 2016

- A. Castellana, A. Carullo, A. Astolfi, *Variabilità intra e inter-soggetto della misura del livello di pressione sonora in materiale vocale acquisito con un analizzatore vocale e un fonometro*, Congresso Nazionale GMEE – Benevento (Italy), 19-21 September 2016

- A. Castellana, A. Carullo, S. Corbellini, A. Astolfi, M. Spadola Bisetti and J. Colombini, *Cepstral Peak Prominence Smoothed distribution as discriminator of vocal health in sustained vowel*, in Proc. IEEE I2MTC, Torino (Italy), May 22-25, 2017, pp. 552-557.

- A. Astolfi, A. Carullo, A. Castellana, G.E. Puglisi, M. Spadola Bisetti, *Monitoraggio della voce per la valutazione della salute vocale e dell'influenza dell'acustica dell'ambiente*, LI Congresso SIFEL, Torino (Italy), 16-17 June 2017

- A. Castellana, A. Carullo, A. Astolfi, *Variabilità intra e inter-soggetto della misura del livello di pressione sonora*, TUTTO MISURE, 2, 95-98 (2017)

- A. Castellana, A. Selamtzis, G. Salvi, A. Carullo, A. Astolfi, *Cepstral and entropy analyses in vowels excerpted from continuous speech of dysphonic and control speakers*, in INTERSPEECH 2017, pp. 1814-1818.

- A. Astolfi, A. Carullo, S. Corbellini, M. Spadola Bisetti, G.E. Puglisi, A. Castellana, G. D'Antonio, L. Pavese L. Shtrepi, A. Peretti, A. Pierobon, J. Griguolo, G. Marcuzzo, G.B. Bartolucci, *Il monitoraggio dei parametri vocali degli insegnanti*, 80 Congresso Nazionale SIMLII – Padova (Italy), 20-22 September 2017

The following abstracts were presented at national and international conferences:

- A. Carullo, A.R. Accornero, A. Astolfi, A. Castellana, L. Pavese, G. Pecorari, G.E. Puglisi, *A low-cost portable vocal analyzer for long-term monitoring and clinical investigation*, 11th International Conference on Advances in Quantitative Laringology and 4th International Occupational Voice Symposium 2015, London (UK), 8-10 April 2015

- M. Spadola Bisetti, J. Colombini, A. Accornero, A. Castellana, G.E. Puglisi, A. Carullo, S. Corbellini, A. Astolfi, R. Albera, *Progettazione e validazione di un "Holter Vocale": dispositivo indossabile per il monitoraggio continuo della voce*, 35° Congresso Nazionale SIAF, Milano (Italy), 17-19 December 2015

- A. Castellana, A. Carullo, S. Corbellini, U. Fugiglando, J. Colombini, M. Spadola Bisetti, A. Astolfi, G.E. Puglisi, *CPPS Distributional Shape and Parameters as Effective Tools to Discriminate Dysphonic and Healthy Voice*, 45th Annual Voice Symposium: Care of the Professional Voice – Philadelphia (Pennsylvania, USA), 30 May-3 June 2016

- A. Astolfi, G. Calosso, G.E. Puglisi, A. Castellana, A. Carullo, *Professional voice use in high school classrooms: relationships between classroom acoustics and voice parameters of teachers at the beginning and at the end of a school year*, 45th Annual Voice Symposium: Care of the Professional Voice – Philadelphia (Pennsylvania, USA), 30 May-3 June 2016

- A. Astolfi, G.E. Puglisi, L. Shtrepi, A. Carullo, S. Corbellini, A. Castellana, A. Accornero, M. Spadola Bisetti, *Monitoring voice over a long period*, 5th Symposium of the Finnish Society of Voice Ergonomics – Helsinki (Finland), 9 September 2016

- A. Astolfi, G.E. Puglisi, L. Shtrepi, A. Carullo, S. Corbellini, A. Castellana, A. Accornero, M. Spadola Bisetti, *A longitudinal study on vocal behaviour of teachers in classrooms and relationships with classroom acoustics*, 5th Symposium of the Finnish Society of Voice Ergonomics – Helsinki (Finland), 9 September 2016

- A. Castellana, G.E. Puglisi, G. Calosso, A. Accornero, L.C.C. Cutiva, F. Fanari, F. Pellerey, A. Carullo, A. Astolfi, *One-year longitudinal study on teachers' voice parameters in secondary-school classrooms: Relationships with voice quality assessed by perceptual analysis and voice objective measures*, Acoustics '17 Boston – Boston (Massachusetts, USA), 25-29 June 2017

- A. Castellana, A. Carullo, A. Astolfi, G. E. Puglisi, U. Fugiglando, *Speech Sound Pressure Level distributions and their descriptive statistics in successive readings for reliable voice monitoring*, Acoustics '17 Boston – Boston (Massachusetts, USA), 25-29 June 2017

- A. Astolfi, G.E. Puglisi, L. Shtrepi, A. Carullo, S. Corbellini, G. D'Antonio, A. Castellana, A. Accornero, M. Spadola Bisetti, A. Peretti, G. Marcuzzo, A. Pierobon, G.B. Bertolucci, *Monitoring of voice over a long period with smartphone applications and contact microphone*, Acoustics '17 Boston – Boston (Massachusetts, USA), 25-29 June 2017

During these three years I also have had the opportunity to collaborate in studies in the electronic measurement field, which have led to the following publications:

- A. Carullo, A. Castellana, A. Vallan, A. Ciocia, F. Spertino, *Definition and preliminary results of degradation tests for photovoltaic modules*, 21 IMEKO TC4 International Symposium – Budapest (Hungary), 7-9 September 2016

- A. Carullo, A. Castellana, A. Vallan, A. Ciocia, F. Spertino, *Uncertainty issues in the experimental assessment of degradation rate of power ratings in photovoltaic modules*, Measurement, 111, 432-440 (2017)

- A. Carullo, A. Castellana, A. Vallan, A. Ciocia. F. Spertino, *Degradation rate of eight photovoltaic plants: results during six years of continuous monitoring*, 22nd IMEKO TC4 International Symposium & 20th International Workshop on ADC Modelling and Testing, Iasi (Romania), 14-15 September 2017

- A. Carullo, A. Castellana, A. Vallan, *Degrado delle prestazioni di otto impianti fotovoltaici durante sei anni di monitoraggio continuo*, I Forum Nazionale delle Misure, Modena (Italy), 13-16 September 2017

Chapter 2

Secondary school teachers' vocal behaviour and influence of classroom acoustics in a school year longitudinal study

This chapter partially reports material from:

1. G. Calosso, G.E. Puglisi, A. Astolfi, A. Castellana, A. Carullo, F. Pellerey, *A 1-school year longitudinal study of secondary school teachers' voice parameters and influence of classroom acoustics*, The Journal of the Acoustical Society of America, 142 (2), 1055-66 (2017)

This chapter describes a one-school year longitudinal study of teachers' vocal behaviour during working hours. As highlighted in Chapter 1 (paragraph 1.2), in the existing literature several works deal with in-field long-term monitorings of teachers' voice use, but there is a lack of longitudinal studies that assess voice parameters modifications accounting for room acoustic-related factors (reverberation and background noise level in classrooms).

Particularly, the objectives of the work were as follow:

1. to investigate the changes in the teacher's voice production, and thus identify the risk of vocal dysfunctions of the teachers after one year of working activity;

2. to determine the variation in the noise conditions measured at the beginning and at the end of a school year;

3. to investigate the influence of noise, reverberation, and their combined effect on voice production in order to set up a predictive model that would be able to estimate the speech sound pressure level from the background noise level and the reverberation time.

2.1 Voice monitoring of the teachers

Teachers involved in the present study work in two secondary schools located in the Province of Turin (Italy). A brief description of the schools is shown below, in order to underline the main differences between the two, i.e., their location, their construction period, the classroom dimensions, and the acoustic characteristics. School A was built in the early 1800s, and is located in the city center close to a heavy vehicular traffic road. The façades of the building are made of bricks and large, modular, single-plane windows. The internal spaces are separated by lightweight brick walls and simple wooden doors. The school classrooms differ in volume, which ranges between 180 m^3 and 400 m^3, and therefore also in acoustic characteristics. A total of 34 classrooms have been considered in the present study, and of these, only 14 have absorptive false ceilings. School B is dated back to the second half of the 1900s and it is located in a suburban area, where only quiet roads are present. This building is made of reinforced concrete, prefabricated elements, and double-glazed sliding windows. The internal spaces are separated by light-weight plasterboard walls, and plasterboard doors with ventilation grids. The 13 classrooms considered in this school have all had acoustic treatment in the form of a false absorbing ceiling with volumes ranging between 170 m^3 and 210 m^3. Figure 2.1 shows the buildings of the two schools.

Thirty-one teachers from the two secondary schools were involved at the beginning of the school year (stage 1): 21 in school A, 4 Males (M) and 17 females (F), and 10 in school B, 2 M and 8 F. Twenty-two of them (14 in school A, 2 M and 12 F, and 8 in school B, 2 M and 6 F) also participated at the end of the same school year (stage 2). Their vocal activity was monitored for two to three working days over two weeks in each stage. Their age ranged between 38 and 62 years, with a mean age of 52. Only the 22 teachers who took part in both stages were considered to assess the

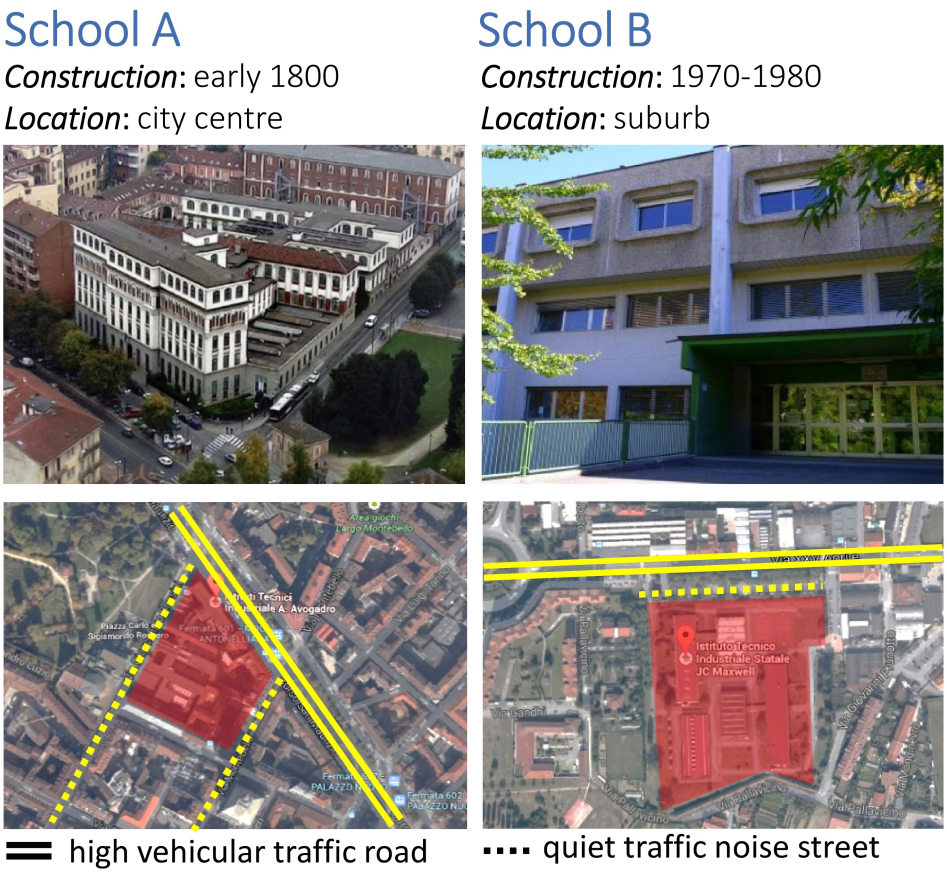

Fig. 2.1 The buildings of the two schools involved in this study.

changes in the voice production over the school year. The teachers who were only monitored during stage 1 were considered to study the relationships between voice parameters, noise, and reverberation at the beginning of the school year. Physical education teachers were excluded from the study since they are subjected to a higher vocal effort than science and humanity teachers. The teachers' vocal activity was monitored using Voice Care [30, 31], which has been described in Chapter 1 (paragraph 1.2). This device is connected to an ECM that is used as a contact microphone to sense the acceleration of the skin due to the vibration of the vocal folds. Such a contact microphone is suitable for long-term monitorings of voice during work activities, since the acquired signal is negligibly affected by background noise [104]. The off-line processing allows the following vocal parameters to be extracted from the recorded signal: the sound pressure level at 1 m in front of the speaker's mouth (SPL_{1m} in dB), the fundamental frequency (F0 in Hz) and the voicing time percentage (Dt% in %). As described in the Introduction chapter, such evaluations are based on the *voiced* and *unvoiced* frame detection through a suitable *rms* voltage threshold, where each frame is 30 ms long. In particular, a proper Matlab script has been implemented to obtain SPL_{1m} and F0 occurrence histograms from *voiced frames* with a bin resolution of 1 dB. Mean, mode, and standard deviation values have been calculated from such histograms, obtaining $SPL_{mean,1m}$, $SPL_{mode,1m}$, and $SPL_{sd,1m}$, and $F0_{mean}$, $F0_{mode}$, $F0_{sd}$, respectively. Also the equivalent SPL at 1 m from the speaker's mouth ($SPL_{eq,1m}$) has been estimated, which expresses the speaker's vocal effort according to the ANSI S3.5–1997 standard [61]. $SPL_{eq,1m}$ has been calculated as the average of the voiced energy over all the frames, including the unvoiced ones, whose energy is set to zero, according to Švec *et al.* [18] as follows:

$$SPL_{eq} = 10log\left(\frac{1}{N}\sum_{i=1}^{N} n_i \cdot 10^{\frac{SPL_i}{10}}\right) \tag{2.1}$$

where N is the total number of frames in the analyzed speech and n is equal to 0 for the unvoiced frames and 1 for the voiced frames.

Males and females were considered together for SPL and Dt% values, whereas the two genders were considered separately for the F0 statistics, but then only the female subject values were kept since male subjects were only 6. The voice monitoring of the teachers consisted of acquiring voice parameters under two conditions:

1. in conversational conditions, i.e., conversational voice parameters (CVPs);

2. in occupational conditions, i.e. occupational voice parameters (OVPs);

In order to obtain CVPs, a voice sample was acquired before the teaching activity had started: each teacher was asked to talk for 5 minutes using a conversational pitch of voice, while seating at 1 m of distance in front of the listener in a silent school room with room acoustics similar to the classrooms where lessons took place. In order to obtain OVPs, the vocal activity of the teachers was monitored over several working hours and days. Since a typical lesson period consists of various activities with subsequent changes in the voice use and in the noise conditions, a specific activity, i.e. the plenary lesson, has been chosen to evaluate the OVPs. Plenary lesson was found to be the most frequent activity conducted by the teachers, and it had been monitored for 74% of the total time: during this type of lesson, the teacher generally speaks in front of the class with students listening, and only one person speaks at a time. The duration of the plenary lessons, excluding recreation time, ranged between 45 min and 60 min.

2.2 The classroom acoustics

The acoustic characterization of the classrooms was performed in both the schools in simulated occupied conditions, using absorptive polyester fiber panels that were dimensioned in order to have the same absorptive properties as 23 seated teens. The measurements were carried out in compliance with the BS EN ISO 3382-2 standards,[105] applying the integrated impulse response method. A pair of wooden boards, hinged together to generate impulsive signals and a sweep signal generated by the B & K type 4128 Head and Torso Simulator, HATS, (Nærum, Denmark) were used in school A and B, respectively. The signals, which were measured in two source and four microphone positions, were averaged to obtain the mean spatial values. Moreover, the reverberation time ($T30_{0.25\div2\text{kHz,occ}}$) was averaged in frequency between 0.25 and 2 kHz, according to the DIN18041 German standard [106]. This standard specifies a range of acceptable values that are defined as a function of volume (V), the mode of use of the room and the typical speech spectrum. The measured classrooms have been grouped into seven room types which had similar volumes and reverberation times. Table 2.1 reports mean and standard deviation of volume and reverberation time for each classroom type, and the number of the rooms included in the same type. The volumes of classrooms vary a lot in school A,

Table 2.1 Classroom characteristics: volume and reverberation time ($T30$) measured in occupied conditions. The standard deviation is reported in brackets when repeated measurements were taken. Values in bold indicate the values that are in compliance with the optimal range of the DIN 18041 standard.

Classrooms	School A					School B	
	1.A	2.A	3.A	4.A	5.A	1.B	2.B
Volume (m^3)	180	210	280	320	400	170	210
Number	7	8	7	5	7	9	4
$T30$ (s)	**0.7 (0.18)**	1.1 (0.26)	0.8 (0.22)	1.4 (0.11)	1.6 (0.29)	**0.5 (0.07)**	**0.5 (0.11)**

Fig. 2.2 Left side: a huge volume classroom, with high ceiling and no absorbing surfaces of school A; right side: a classroom with limited volume and absorbing ceiling of school B.

reaching 400 m^3, while school B has only two room sizes, which are smaller than the previous case. Table 2.1 also underlines that all the $T30_{0.25\div2kHz,occ}$ values in school B are equal to 0.5 s, thus complying with the reference value [106], while the same is not in school A ($T30_{0.25\div2kHz,occ}$ ranging between 0.7 and 1.6 s). In Figure 2.2 two classrooms of the schools with opposite acoustics are shown: on the left there is a huge volume classroom, with high ceiling and no absorbing surfaces of school A; on the right, instead, there is a classroom with limited volume and absorbing ceiling of school B.

The background noise activity levels were measured at the same time as the teachers' voice monitoring. The measurements were carried out using a class 1 sound level meter. The sound level meter was positioned close to the teacher's desk, at least 1 m away from any reflecting surface and at 1.2 m from the ground,

in compliance with the ISO 1996 recommendations [107]. During the monitoring periods, the classrooms were occupied by an average number of 23 students.

The background noise level was evaluated as the A-weighted level exceeded for 90% of the considered time (L_{A90} in dB). All the measurements were performed for a time interval of 15 min, during which a researcher was present in the classroom to take note of the different activities and noisy events that occurred during the noise monitoring. In the absence of particularly noisy events, the 15 min records were considered representative of the entire activitycarried out in the same period, which on average lasted an hour. This choice was made on the basis of a previous work by Puglisi *et al.*, [108] in which L_{A90} measured for 15 min was not found to be statistically different from the values obtained over long-time measurements of 4 hours. For school A, which was located close to heavy traffic roads, only the samples of noise acquired in the classrooms that faced onto the courtyards were considered. In such a way, the noise measurements were not affected by the noise from outside, but only by the indoor noise, thus allowing the relationship between measured noise and acoustic treatment of the classrooms to be investigated.

2.3 Analyses

The statistical analyses of data have been carried out with SPSS software (v. 21; SPSS Inc, New York). The Shapiro–Wilk test has been first applied to understand whether the parameters related to the classroom acoustics and teachers' voice are normally distributed. All the calculations have been performed considering a confidence interval of 95% (significance level of 0.05). In the following subsections, a description of the preformed analyses for each goal of this work is reported.

2.3.1 Longitudinal study of the teachers' voice parameters and background activity noise conditions

The two tailed Paired Sample *t*-test has been performed to examine the variation in the voice production of teachers over a school year. The test calculates the difference for each subject within each before-and-after pair of measurements, taking into account the within-subject dependence of the voice parameters. Since a single pair

of values must be attributed to each subject in the Paired Sample *t*-test, the means of the repeated measurements of the voice parameters acquired for each subject during stage 1 and stage 2 were associated to each individual. The teachers in the two schools have been analyzed separately, since the acoustic conditions of the two schools were different, thus having affected the teachers' vocal behaviour in a different way. The Independent Sample *t*-test has been used to assess the differences in the mean values of L_{A90} between the two schools and the variation in L_{A90} between the beginning and the end of the school year. Therefore, the activity background noise measurements have been divided into four different groups:

1. Group 1, L_{A90} measurements in school A at stage 1.
2. Group 2, L_{A90} measurements in school A at stage 2.
3. Group 3, L_{A90} measurements in school B at stage 1.
4. Group 4, L_{A90} measurements in school B at stage 2.

2.3.2 Relationships between the classroom acoustics and the teachers' voice parameters

The relationships between the average changing pattern of OVPs and the acoustic conditions inside the classrooms (background noise level and reverberation time) have been first assessed using the single variable regression analysis. This type of analysis allows results to be compared with previous studies that had investigated the effect of the acoustic conditions of the classroom on the teachers' vocal behavior [48, 49, 109]. OVP values have been grouped together and averaged on the basis of the independent variable classes for $T30_{0.25\div2\text{kHz,occ}}$. The classes have been defined using a just noticeable difference equal to 5%, according to BS EN ISO 3382–2:2008, [105], in order to obtain well defined and robust groups of data. Since the analysis has been ran both at the beginning and at the end of the school year, it has been possible to establish whether the relationships between the different parameters are the same in the two periods of the school year. Although this analysis is the most frequently used one, it suffers from some limitations:

1. Noise and reverberation are both present during teaching activities, but the current studies analyze each parameter with single regression models, with-

out considering the combined effect of the two parameters through multiple regression models.

2. Several studies have recognized that a considerable amount of the variance in the voice parameters is due to inter-subject differences, which is also called within-subject dependence. This dependence is not taken into account in a simple regression analysis.

Therefore, the combined effect of reverberation time and background noise level on the teachers' voice parameters has been evaluated by means of linear mixed effects model (LME) [110], which takes into account the within-subject dependence of the teachers' voice parameters and provides a general, flexible approach that can be used to model correlated data obtained from repeated measurements. The term "mixed" refers to the use of both fixed and random effects in the same analysis. Fixed effects are explanatory variables that affect the average response of all the outcomes. Random effects instead have levels that are thought of as random selections from a much larger set of levels. In the present study, reverberation time and background noise level are the fixed effects, which are considered as continuous variables, while subjects are the random effects. The basic idea is that the fixed effects provide estimates of the average rate of change of the population, while the random effect parameters present the general variability among the subjects, which is not achievable with a simple regression analysis. In general, the mixed effect model, for a response variable Y, which depends on the *i-th* subject, can be presented as follow:

$$Yi = (a + \alpha_{\mathrm{i}}) + (b_1 \cdot X_1) + (b_2 \cdot X_2) + ... + \varepsilon_{\mathrm{i}} \tag{2.2}$$

where X_1, X_2,... are the fixed effect parameters; b_1, b_2, are the coefficients of the fixed effect parameters; a is the fixed part of the intercept; α_{i} is the random part of the intercept; and ε_{i} is the residual or unexplained variation, which is also considered a random effect. The standard deviations of random effects α_{i} and ε_{i} are denoted as σ_α and σ_ε, respectively. Moreover, σ_α indicates the general variability between subjects in equal surrounding conditions, while σ_ε represents the variability of individual value around the individual regression line. This analysis has been carried out considering all the measurements acquired in the two stages, since the main objective is to find a general model to estimates the speaker's voice parameters from the acoustic conditions of the classroom, regardless of the period of the school

year. LME was fitted by means of Restricted Maximum Likelihood. The backward form of the stepwise analysis was applied to establish the importance of considering the interaction between L_{A90} and $T30$.

2.4 Results and discussion

2.4.1 Teachers' voice parameters

Table 2.3 shows the mean and the standard deviation of the mean (standard error, SE) of the OVPs and CPVs that changed significantly (p-value < 0.05) over the school year. In the conversational condition, $SPL_{mean,1m}$ increased at the end of the school year in both the schools: it rose by 4.6 dB (SE = 1.0 dB) and 3.0 dB (SE = 0.7 dB) in school A and B, respectively. As far as the OVPs are concerned, significant variations between the two stages have been only observed in school A, that is, the school with the higher noise and reverberation time values. $SPL_{mean,1m}$ increased on average by 2.3 dB (SE = 1.0 dB), whereas $D_{t\%}$ and $F0_{sd}$ decreased by 10.3% (SE = 1.9%) and 4.5 Hz (SE = 1.4 Hz), respectively. Table 2.4 shows the mean $SPL_{eq,1m}$ and the corresponding type of vocal effort for each school, stage, and conversational/occupational voice condition. In agreement with the ANSI S3.5 standard [61], it has been found that the mean vocal effort ranges from "normal" ($SPL_{eq,1m}$ < 65 dB) to "raised" (65 dB $\leq SPL_{eq,1m}$ < 71 dB) in the case of conversational voicing, and from "normal" to "loud" (71 dB $\leq SPL_{eq,1m}$ < 78 dB) for occupational voicing. No significant variations in the mean value of $SPL_{eq,1m}$ have been observed between the two stages. Figure 2.3 shows the percentages of occurrence of the different vocal effort ratings observed over the year in the two schools. It can be seen that, at the beginning of the school year, the most frequently occurring vocal effort is of "shout" type ($SPL_{eq,1m} \geq 78$ dB), whereas "loud" occurred more frequently at the end of the year.

The increase in the conversational $SPL_{mean,1m}$ at the end of the school year in both schools for the same noise condition makes it possible to assume that the background noise level is not one of the causes of the increase in the conversational level of voice at the end of the school year. Therefore, the increase in the conversational level of voice at the end of the school year could be explained as a consequence of the need to use high voice levels during working activities that make teachers use a higher

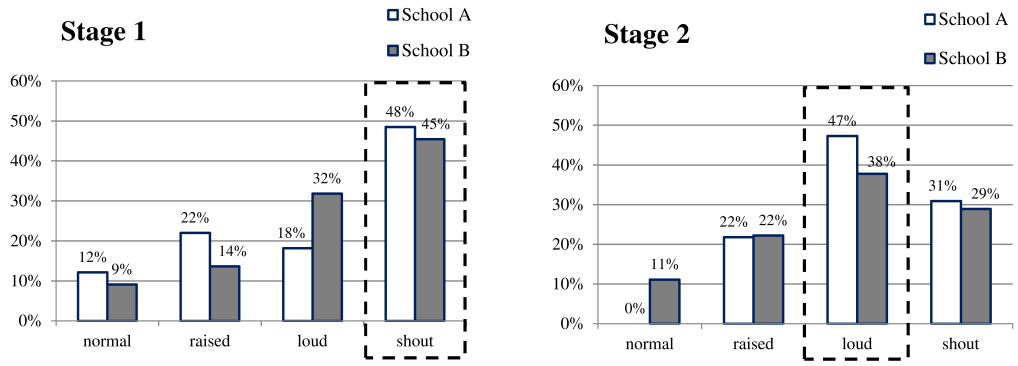

Fig. 2.3 Incidence of the different vocal effort ratings measured at the beginning (stage 1) and at the end (stage 2) of the school year in the two schools.

level of voice during non-occupational activities. As far as the OVPs are concerned, the increase in $SPL_{\mathrm{mean,1m}}$ of 2.3 dB in the school with the higher reverberation time and noise level values is in accordance with previous studies by Laukkanen *et al.* [38], who found an increase in the teacher's sound pressure level during one day at work, and described it as most likely being the result of an adaptation to prolonged voice use. Moreover, the increase of 2.3 dB in $SPL_{\mathrm{mean,1m}}$ is greater than the uncertainty contribution for the mean value of SPL_{mean} that is equal to 0.6 dB according to Castellana *et al.* [60]. They obtained an inter-speaker variability of 2.8 dB for SPL_{mean} estimated by the Voice Care in repeating readings, and we divided such variability to the root square of the number of teachers involved in our study. In this way the uncertainty contribution of SPL_{mean} in the group of teachers was defined, and a proper comparison of the measures has been possible.

Concerning the fundamental frequency, the results do not confirm what was pointed out by Rantala *et al.* [41], Laukkanen *et al.* [38] and Hunter and Titze [39] who studied teachers' vocal behavior during working hours, since no significant variations in the mean F0 were found in our study. Only the $F0_{\mathrm{sd}}$ was found to decrease by 4.5 Hz at the end of the school year, and only for the teachers from school A. However, the Rantala *et al.* [41], Laukkanen *et al.* [38] and Hunter and Titze [39] studies assessed the variation in the voice parameters over one working day, and it is therefore reasonable to presume that different results could be obtained when vocal behaviour is analyzed over a longer period.

The teachers' vocal effort has been evaluated in terms of $SPL_{\mathrm{eq,1m}}$ in order to comply with the ANSI S3.5 standard. No significant variations were observed in

the mean value of $SPL_{eq,1m}$ between the two stages. Nevertheless, it was found that, at the end of the school year, the most recurring vocal effort was "loud" (71 dB $\leq SPL_{eq,1m} <$ 78 dB), whereas it was "shout" at the beginning ($SPL_{eq,1m} \geq$ 78 dB). These values indicate that the teachers exerted an excessive vocal effort in both periods of the school year, and this constitutes a high risk factor for their vocal health. The decrease in the number of teachers with a "shout" vocal effort at stage 2 could mean that the subjective feeling of fatigue increases at the end of the school year, and as a result, teachers tend to decrease their vocal effort as a prevention strategy. This result is not in contrast with the increase in $SPL_{mean,1m}$ at the end of the school year. In fact, $SPL_{eq,1m}$ is the average of voiced energy over all the frames, including the unvoiced ones, whereas $SPL_{mean,1m}$ is calculated by excluding the unvoiced voice frames. Therefore, the increase in the $SPL_{mean,1m}$ and the decrease in the vocal effort ($SPL_{eq,1m}$) could reveal that teachers tend to speak with higher voice levels, but at the same time reduce the number of voiced frames. This can be confirmed in school A considering the increase in the $SPL_{mean,1m}$ and, at the same time, the decrease in the Dt% at the end of the school year. The voicing time percentage values measured during the plenary lessons in both stages are significantly higher than those found in the Hunter and Titze [39] and Bottalico and Astolfi studies [48], which were between 23 and 30%. One reason for the lower values in the previous studies could be due to the fact that the vocal data had been acquired over all the working hours, during different types of activities, and also with long periods of silence between lessons.

2.4.2 Teaching activity and background noise level

Fifty-five noise monitorings were performed in the plenary lessons, 29 at the beginning of the school year and 26 at the end. Table 2.2 shows the arithmetically averaged measured L_{A90} values for the two schools and the two stages of the school year. The unoccupied L_{A90} levels, measured in the classrooms where the conversational samples of voice had been acquired, were 42.3 dB (standard deviation = 0.3 dB) and 46.1 dB (standard deviation = 0.2 dB) at the beginning of the school year, and 42.4 dB (standard deviation 0.2 dB) and 45.8 dB (standard deviation = 0.5) at the end of the school year in schools A and B, respectively. Table 2.2 also shows that during the teaching activities, L_{A90} was not significantly different in the two schools at the beginning of the year, while at the end of the school year it was significantly higher in school A, which is characterized by higher values of reverberation time.

Table 2.2 Mean values of a certain number of measurements, No., of the background noise level (L_{A90}) in the two schools during the two stages. Significant differences among the mean values of L_{A90} in the two stages (p-value<0.05) are identified with symbol *. Values in bold indicate significant different means of L_{A90} between the two schools during the same stage (p-value<0.05). The standard deviation of the mean (standard error, SE) is reported in brackets.

	L_{A90}							
	Stage 1			Stage 2			Difference (2-1)	
	No.	Mean	SE	No.	Mean	SE	Mean	SE
School A	13	48.0	1.0	14	**59.0**	1.0	+11.0*	1.4
School B	16	46.9	0.9	12	**53.5**	0.9	+6.6*	1.3

Moreover, significant L_{A90} increases of 11.0 dB and 6.6 dB have been observed at the end of the school year in schools A and B, respectively.

The arithmetically averaged L_{A90} values for the two schools and the two stages of the school year were in the range between 46.9 dB (SE 0.9 dB) (school B, stage 1) and 59.0 dB (SE 1.0 dB) (school A, stage 2). These results are in close agreement with the values measured by Shield *et al.* [50], who found L_{A90} values of between 38 and 63 dB in secondary schools during teaching activities. The L_{A90} value was not significantly different in the two schools at the beginning of the year, while it was significantly higher in school A, which is characterized by higher reverberation time values, at the end of the school year. The significant increase of 11.0 and 6.6 dB observed at the end of the school year in schools A and B, respectively, reveals that, after one year of exposition to high noise levels, the students tend to make more noise. On the basis of the results of the noise monitoring method, the authors believe that the students' behaviour is the main reason for the noise increase, since the noise measurements were taken in classrooms that were not affected by noise from outside, e.g., heavy vehicular traffic. Furthermore, the noise measurements were acquired during the same type of lesson in order to consider similar types of activity noise. Therefore, the feeling of fatigue at the end of the school year could be considered one of the reasons for the noise increase. Furthermore, it cannot be excluded that the feeling of fatigue supports the Lombard effect, that is, the tendency of students to increase their level of voice in noisy conditions.

2.4.3 Classroom acoustics and teachers' voice parameters

Voice and noise

As far as the influence of the background noise level, L_{A90}, on the OVPs is concerned, an increase in $SPL_{mean,1m}$ and in $F0_{mean}$ has been observed as L_{A90} increased (Figure 2.4). In particular, 0.4 and 0.2 dB increases in $SPL_{mean,1m}$ per each 1 dB increase in L_{A90} have been found during stage 1 and stage 2, respectively. Increases of 2.4 and 2.7 Hz of $F0_{mean}$ per each 1 dB of increase in the background noise level have been found at the beginning and at the end of the year, respectively. Furthermore, Figure 2.4 shows that the regression line between $F0_{mean}$ values and L_{A90} related to stage 2 (R^2=0.82 and p-value<0.05) is below the regression line related to stage 1 (R^2=0.37 and p-value<0.05).

As far as the relationship between the background noise level L_{A90} and OVPs is concerned, both an increase in $SPL_{mean,1m}$ and in $F0_{mean}$ with the increase of L_{A90} was observed. It is interesting to note that the increase in $SPL_{mean,1m}$ with the noise at the beginning of the school year (0.4 dB/dB) was higher than that observed at the end of the year (0.2 dB/dB). At the beginning of the school year, the slope confirms a Lombard effect, which is in agreement with the results by Lane and Tranel [46], Bottalico and Astolfi [48], Sato and Bradley [49], and Durup *et al.* [109], who found linear relationships between SPL and the noise level with slopes ranging from between 0.3 and 0.7 dB/dB.

At the end of the school year, SPL values were higher than those measured at the beginning of the school year, thus the teachers generally tended to speak with higher levels of voice. However, their level of voice did not significantly change with noise(only 0.2 dB/dB), not supporting a Lombard effect. Although the different slopes found at the beginning and at the end of the school year confirm the great variability of the Lombard effect across speakers and noise levels stated in literature [46, 47, 111], this result may also indicate that speakers at the end of the year have difficulty maintaining the same increase of their level of voice when the background noise level becomes very high, i.e., higher than 50 dB(A) as in the present study.

The increases of $F0_{mean}$ with the beackground noise in both the stages corroborate the results of Bottalico and Astolfi [48], who found a trend of 1.0 Hz/dB. Moreover, the regression line between $F0_{mean}$ and L_{A90} in the stage 2 is below the

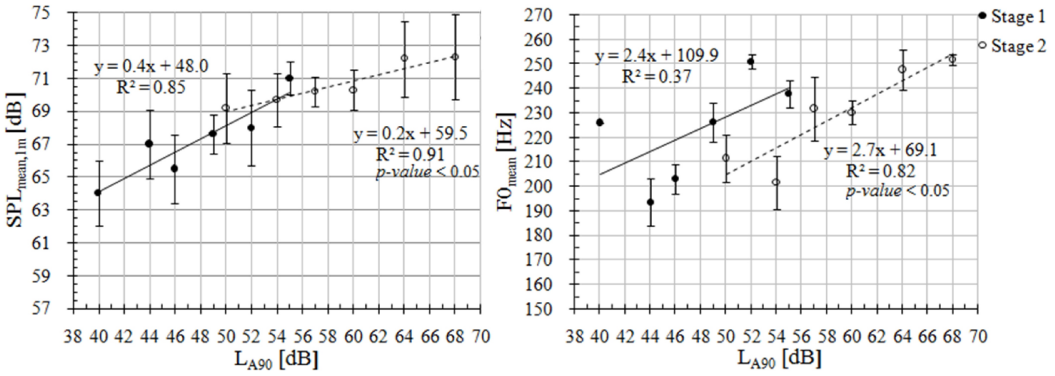

Fig. 2.4 Best-fit linear regressions between the occupational voice parameters ($SPL_{\mathrm{mean,1m}}$ and $F0_{\mathrm{mean}}$) and background noise level (L_{A90}) monitored during stage 1 and stage 2. Each experimental datum on the graph represents the mean value of an average of 5 pairs. The error bars refer to the standard deviation of the means (standard error, SE).

regression line related to stage 1, thus indicating that after a year of working activity the subjects showed a lower F0, even in the same noise conditions.

Noise and reverberation

Figure 2.5 shows the best-fit regression line for the mean L_{A90} vs $T30_{0.25\div 2\mathrm{kHz,occ}}$. The graph indicates an average rate of 5 dB/s in both periods of the year. The regression line related to stage 2 (R^2=0.46 and p-value<0.05) moves upwards, with respect to that of stage 1 (R^2=0.58 and p-value<0.05), thus indicating an increase in the noise level at the end of the school year, as seen in Sec. 2.4.2.

As far as the relationship between L_{A90} and $T30_{0.25\div 2\mathrm{kHz,occ}}$ is concerned, the noise increase with reverberation time indicates that the background noise produced by the students was affected by the reflections present in the classrooms in both periods of the school year. This result confirms the strong linear relationship found by Puglisi *et al.* [21], who studied classroom acoustic conditions and the voice use of primary school teachers.

Voice and reverberation

Figure 2.6 describes the relationships between the mean values of $SPL_{\mathrm{mean,1m}}$ and $T30_{0.25\div 2\mathrm{kHz,occ}}$: teachers adjust their voice levels with the reverberation time in

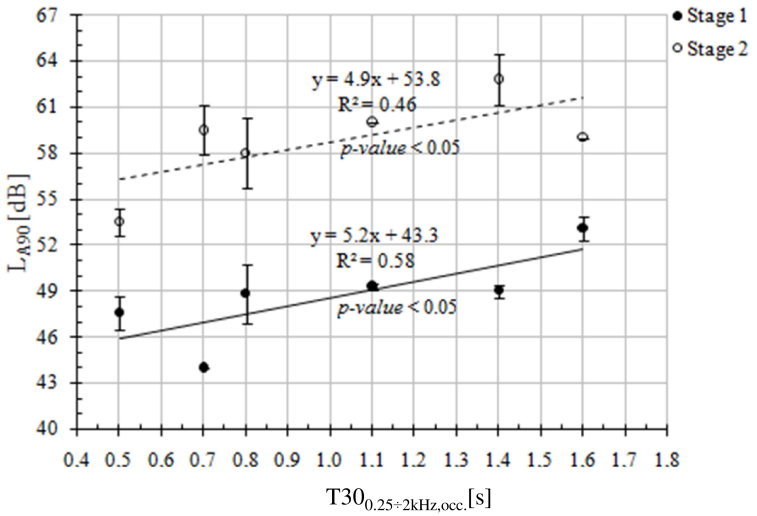

Fig. 2.5 Best-fit linear regressions of the background activity noise levels during the working hours (L_{A90}) and the mid-frequency reverberation time in occupied conditions ($T30_{0.25\div2kHz,occ}$) measured during stage 1 and stage 2. Each experimental datum in the graph represents the mean value of an average of 10 pairs. The error bars refer to the standard deviation of the means (standard error, SE).

classrooms according to a quadratic regression in both the stages. The minimum values of such curves correspond to a $T30_{0.25\div2kHz,occ}$ of 0.83 and 0.77 s for stage 1 and stage 2, respectively. It should be noted that the regression curve related to stage 2 (R^2=0.94) is above the curve related to stage 1 (R^2=0.63 and p-value<0.05), thus indicating an increase in $SPL_{mean,1m}$ at the end of the school year, as shown in Sec. 2.4.1.

The two best-fit regression curves between $SPL_{mean,1m}$ and $T30_{0.25\div2kHz,occ}$, with minimum values at 0.83 and 0.77 s of $T30_{0.25\div2kHz,occ}$ at stage 1 and stage 2, respectively, indicate that there was an optimal degree of reverberation that supported the speaker's voice. These results confirm the results of a previous study by Bottalico and Astolfi [48] in primary schools, where a quadratic regression curve and an optimal 0.8 s value of mid-frequency reverberation time were found for occupied classrooms. These minimum values also corroborate the results of a recent study by Puglisi *et al.* [21], where a range between 0.6 and 1.0 s of $T30_{0.25\div2kHz,occ}$ was found to minimize $SPL_{mean,1m}$ and maximize the vocal comfort of primary school teachers in fully occupied classrooms with volumes of 200 m^3. Furthermore, the quadratic regression curves are partially in agreement with the results of the studies

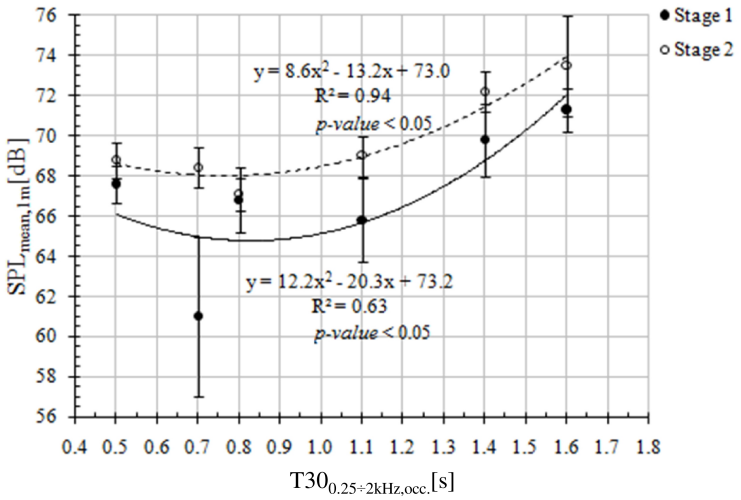

Fig. 2.6 Best-fit quadratic regression curves of the vocal efforts of the teachers ($SPL_{\mathrm{mean,1m}}$) and the mid-frequency reverberation times in occupied conditions ($T30_{0.25\div 2\mathrm{kHz,occ}}$) during stage 1 and stage 2. Each experimental datum in the graph represents the mean value of an average of 10 pairs. The error bars refer to the standard deviation of the means (standard error, SE).

of Brunskog *et al.* [55] and Pelegrín-García *et al.* [56], which showed a tendency of the speakers to lower their voice levels as the reverberation increased. In the present study, this tendency has only emerged for lower values of $T30_{0.25\div 2\mathrm{kHz,occ}}$ than the optimal value, and it is therefore likely that the room does not provide sufficient support to the speaker's voice. Instead, when the reverberation is sufficient to support the speaker's voice, it is possible that a high reverberation produces a higher level of noise, which in turn induces the teachers to raise their voice levels and that confirms the Lombard effect. Optimal reverberation time values in classrooms should also consider speech intelligibility and the acoustic comfort for students, even lots of variability and lack of consistency in terms of grade of students and condition of occupancy is shown in literature and reference standards. Nijs and Rychtarikova [112] indicate an average mid-frequency reverberation time not higher than 0.6 s to preserve speech intelligibility in occupied classrooms of about 200 m^3. Yang and Bradley [113] found an optimal mid frequency reverberation range between 0.3 and 0.9 s for good speech intelligibility in primary school occupied classrooms. As far as current standards on classroom acoustics are concerned, only the BB93 [114] indicates optimal values of reverberation time for secondary schools that

are lower than 0.8 s in unoccupied conditions, averaged in frequency between 0.5 and 2.0 kHz. In France [115], reverberation times up to 0.8 s in unoccupied secondary school classrooms with volumes of 250 m^3 are allowed. In Italy, the UNI 11367 [116] standard defines an optimal reverberation time of 0.8 s in unoccupied conditions, averaged in frequency between 0.5 and 1.0 kHz. As far as optimal values of reverberation time in occupied conditions are concerned, only the German DIN 18041 [106] indicates an optimal range that is function of the room volume for general teaching activity (see Table 2.1). For volumes between 200 m^3 and 400 m^3, the optimal range is $0.6 \div 0.7$ s.

One should note that the dependence of the vocal effort on the total equivalent absorption area in classrooms has not been taken into account, since the aim of this study was the correlation between vocal effort/load and reverberation time. The reason is that the classrooms investigated in the study had similar volumes (the largest room is about two times the volume of the smallest one) and the absorption area was very well correlated with reverberation time, with no high difference among them (the largest absorption area has been found to be about three times the lowest one). However, further longitudinal in-field study that involves different combinations of room volume and absorption area could be useful to provide guidelines for a room acoustic design conceived to optimize vocal comfort and speech intelligibility.

Table 2.3 Paired sample *t*-test of OVPs and CVPs for the two stages of the school year. Significant differences between the two stages (p-value < 0.05) are shown in bold. SE indicates the standard deviation of the mean (standard error) and *df* the degrees of freedom.

		School A (14 subjects)							School B (8 subjects)						
		Stage 1		Stage 2		Difference (2-1)			Stage 1		Stage 2		Difference (2-1)		
	Voice parameter	Mean	SE	Mean	SE	Mean	SE	*df*	Mean	SE	Mean	SE	Mean	SE	*df*
OVPs	$SPL_{\text{mean,1m}}$ (dB)	68.3	1.1	70.6	1.0	**+2.3**	1.0	14	68.4	0.8	68.3	1.3	-0.1	1.5	8
	$D_{\text{t\%}}$ (%)	50.7	2.5	40.4	2.2	**-10.3**	1.9	14	37.4	3.1	38.1	4.1	+0.7	3.4	8
	$F0_{\text{sd}}$ (Hz)	61.7	3.5	57.2	3.5	**-4.5**	1.4	4	58.4	3.7	57.3	2.7	-1.1	2.8	5
CVPs	$SPL_{\text{mean,1m}}$ (dB)	60.4	1.6	65.0	1.1	**+4.6**	1.0	14	58.4	0.7	61.4	0.7	**+3.0**	0.7	8
	$SPL_{\text{mode,1m}}$ (dB)	61.0	1.7	65.5	1.4	**+4.5**	1.2	14	58.7	1.2	60.7	1.5	+2.0	1.1	8

Table 2.4 Mean value and standard deviation of the mean (standard error, SE) of the equivalent sound pressure level at 1 m from the speaker's mouth ($SPL_{\text{eq,1m}}$) and classification of the teachers' vocal effort (VE) during occupational voice use (O) and during conversational voice use (C) for the two stages, according to the ANSI S3.5 standard.

	School A (14 subjects)						School B (8 subjects)					
	Stage 1			Stage 2			Stage 1			Stage 2		
	$SPL_{\text{eq,1m}}$ (dB)			$SPL_{\text{eq,1m}}$ (dB)			$SPL_{\text{eq,1m}}$ (dB)			$SPL_{\text{eq,1m}}$ (dB)		
	Mean	SE	VE	Mean	SE	VE	Mean	SE	VE	Mean	SE	VE
C	64.4	2.6	Normal	65.9	1.9	Raised	62.0	0.8	Normal	61.1	1.1	Normal
O	71.7	1.3	Loud	72.6	1.5	Loud	72.7	1.4	Loud	70.8	1.4	Raised

Table 2.5 *P*-values of the two models tested using a linear mixed effects analysis. Model 1 includes the principal effects and interactions between background noise level (L_{A90}) and reverberation time ($T30_{0.25\div2kHz,occ}$) on the sound pressure level of the speaker ($SPL_{mean,1m}$), while model 2 only includes the principal effects.

Model	Dependent variable	Fixed effects	Random effects	*p*-values
1	$SPL_{mean,1m}$	L_{A90} $T30_{0.25\div2kHz,occ}$ with interaction	Subjects	L_{A90}: 0.001 $T30_{0.25\div2kHz,occ}$: 0.036 $L_{A90}\cdot T30_{0.25\div2kHz,occ}$: 0.044
2	$SPL_{mean,1m}$	L_{A90} $T30_{0.25\div2kHz,occ}$ no interaction	Subjects	L_{A90}: 0.001 $T30_{0.25\div2kHz,occ}$: 0.501

Interaction between voice, noise, and reverberation

Table 2.5 shows the *p*-values of the two models, one with the coefficient of interaction L_{A90}*T30 (model 1) and one without (model 2). Since model 1 has the smallest *p*-values of all the parameters, it would appear that model 1 is better. The multiple mixed effects model, in which the two variables are used and their interaction is considered, results in

$$SPL_{mean,1m} = 31.8 + \alpha_i + 0.7\cdot L_{A90} + 21.6\cdot T30 - 0.4\cdot L_{A90}\cdot T30 + \varepsilon_i \qquad (2.3)$$

Where σ_α is equal to 3.5 dB, which indicates the general variability of SPL between subjects in equal surrounding conditions, and σ_ε is equal to 3.0 dB, which represents the variability of the SPL around the individual regression line for each subject. The linear regression (R^2=0.76) between the measured values versus the predicted values is shown in Figure 2.7. The standard deviation of the residuals is 2.5 dB, which means that 95% of the sample presents residual values that vary from -5 dB to +5 dB.

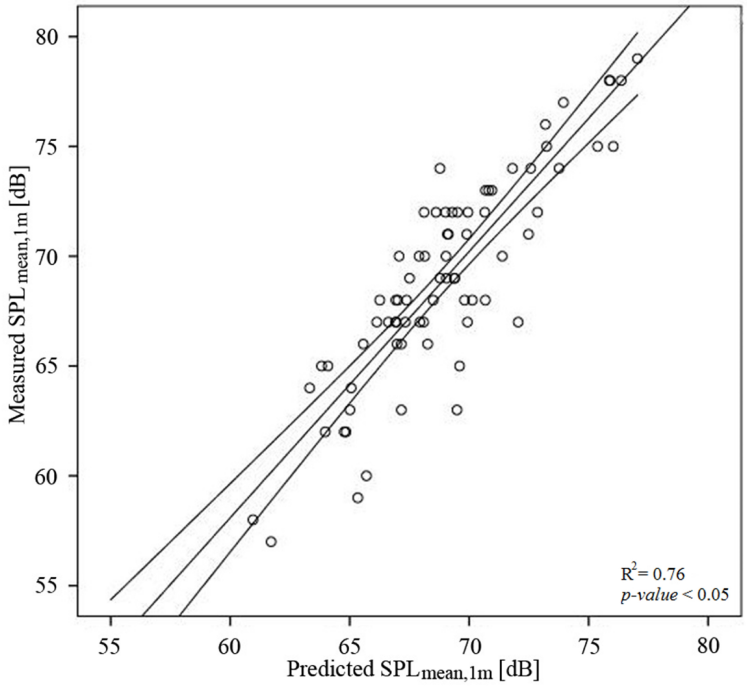

Fig. 2.7 Scatter plot and best-fit linear regression (R^2=0.76) of the measured versus predicted values of $SPL_{\mathrm{mean,1m}}$. The predicted values were estimated using the linear mixed effect models. The solid line shows the linear regression. The curved lines indicate a 95% confidence interval based on the average expected from the regression line.

Chapter 3

Speech sound pressure level distributions and their variability across repeated readings using different devices

This chapter partially reports material from:

1. A. Castellana, A. Carullo, A. Astolfi, G. E. Puglisi, and U. Fugiglando, *Intra-speaker and inter-speaker variability in speech sound pressure level across repeated readings*, The Journal of the Acoustical Society of America, 141(4), 2353–2363 (2017).

This chapter deepens a specific aspect of the uncertainty evaluation of speech Sound Pressure Level measure, that is, the SPL variability due to the voice source unpredictability. In the existing literature, many studies belonging to different fields that are related to speech have evaluated the vocal intensity in terms of statistics or other descriptors of SPL, e.g. the mean, mode or equivalent SPL. On one hand, such investigations are often related to speakers' vocal health: among them, some researchers investigated mean SPL in clinics recordings of a group of patients with vocal nodules and of a control group [117, 118]; others used mean SPL as a descriptor of the effects on vocal function of voice therapy [119]; furthermore, increased average vocal intensity has been related to the presence of vocal fold lesions [120] and the deterioration of vocal fold epithelium [121]. On the other hand, recent investigations deal with the characterization of subject's typical vocal

behavior by means of descriptive statistics of SPL distributions [97, 98], which are one of the output of the developed devices for long-term vocal monitorings, as mentioned in paragraph 1.1.2, Chapter 1. In-field voice monitorings have been used to study the vocal effort of voice professionals, which is a subjective physiological quantity related to voice production that has been computed by means of SPL [122]. In particular vocal effort has been largerly monitored on teachers, since they are one of the categories that are most affected by voice abuse [42, 48]. In Chapter 2, where a one-school-year monitoring of secondary school teachers is described, a wide list of previous investigations is provided. The mean SPL, mode SPL and the equivalent SPL, which is the time-weighted average of SPLs, are the most used SPL parameters for the investigation of occupational vocal risk [123–125, 21]. Moreover, other studies using SPL descriptors are those that have estimated speech SPLs in order to investigate speech modifications due to different room acoustics [126] or noise conditions [56].

In summary, in the existing literature researchers often deal with comparisons between different speech situations, e.g. evaluating speech SPL for a subject or a group of subjects in the case of clinical studies of voice disorders, in the case of occupational evaluation of vocal effort, and in the case of speech modifications due to different acoustics conditions. However, the reported results do not usually take into account the uncertainty contribution related to the repeatability of the subjects involved in the experiments. With the aim to give preliminary results on the spread of repeated measures of SPL of the same subject and of a group of speakers, this study investigates intra-speaker and inter-speaker variability of SPL in continuous speech under repeatability conditions.

Many studies investigated the variability of SPL parameters, focusing on the possible specific causes that generate speech modifications. As summarized by Cooke *et al.* [127], the characteristics related to the addressed listener and to the environment are the dominant situational factors influencing speech production. The effect of the acoustical environment on SPL produced by talkers at different communication distances has been examined in several studies (see Pelegrin-Garcia *et al.* [56] for an overview). Within the communication situations where background noise is present, a global increase of speech intensity occurs, leading to the Lombard speech [128]. However, speech level increases are also observed in the case of talker-to-listener distance increase in absence of noise, perhaps as a form of compensation for perceived listener difficulties. Cushing *et al.* [129] determined a baseline reference for typical

vocal effort considering 50 British English subjects talking in anechoic conditions, extending the Pearsons' dataset (summary of results in [130]) and giving SPL at five different positions around the speaker's head, at the distances of 0.5 m and 1 m.

The speaking style [131] (e.g. clear speech or conversational speech) represents another feature that affects speech production. 'Clear speech' designates any kind of hyper-articulated speech that aims at improving speech intelligibility than ordinary, normally articulated conversational speech. Generally, it is uttered at a lower speaking rate than conversational speech. Lastly, many causes of speech modifications can be related to the talker, e.g. voice status, hearing status, age and gender, mood and physical conditions, speaking experience or training. Byrne *et al.* [132] investigated the long-term average speech spectrum of readings at a normal speed and level, recorded in 12 different languages. Small but statistically significant differences have been found among the languages (most of the variations from the average values were lower than 3 dB) and more substantial differences between male and female talkers at the low frequencies, because of the difference in the fundamental frequency ranges.

Other factors of influence on SPL could be related to the speech task [127] (syllable or vowel, read-speech, spontaneous-speech, simulated-speech, task-oriented speech). An earlier study [133] investigated intra-speaker variation of SPL in 6 subjects related to a reference group data of 15 females and 15 males, while repeating the syllable /pæ/ three times and at three different levels of vocal effort. Other reports investigated the variability of the comfortable effort level across experimental sessions. Brown *et al.* [134] recorded 16 subjects during five successive days while producing three times a series of vowels and phrases and they reported results on within day variation, day to day variation and subject to subject variation of vocal intensity, three different aspects of inter-speaker variability of speech SPL. In a successive work, Brown *et al.* [135] assessed the degree of inter-speaker variability of 50 untrained speakers divided into three age groups in a week across utterance types (vowel, reading and speaking) and recording sessions. Garret *et al.* [136] determined the inter-speaker variation of vocal amplitude in three repetitions of connected speech samples acquired from 20 subjects during three time intervals of one day. Corthals *et al.* [137] investigated vocal intensity in running speech collected from 400 subjects with time-weighted sound pressure level estimates, namely equivalent continuous sound levels and percentile levels, which are adequate descriptors for fluctuating sounds. Sihvo *et al.* [138, 139] studied the repeatability

and reproducibility of sound level measurement of the softest and loudest possible phonations at five given pitches and between 45-minute-long readings.

The results of the reported studies highlighted that vocal intensity varies from one experimental session to the next when subjects were asked to speak in a comfortable manner. Furthermore, the above review reveals that speech SPL and other speech parameters that are related to it [133, 140, 141] vary within and across speakers, owing to all the aforementioned causes of speech modifications.

The present study is about the variability of SPL when subjects speak at a comfortable level. Taking in mind all the factors that affect speech production, a proper experimental design has been planned. Environmental effects have not been taken into account as well as the health status of the subjects, since experiments were performed in a semi-anechoic chamber by young healthy speakers. A selected speech material was used during the experiments, which participants read with a normally articulated conversational speech: in this way, both the speech-task and speech-style effects have been eliminated. The method for estimating measurement uncertainties, which is described in the GUM [142], was followed both in the experimental design and in the result processing. Differently from some above-mentioned studies that focused on the inter-speaker variability of SPL at a comfortable vocal effort, our tests were performed in a day and repeated measurements for each speaker were done consequently, in a 15-minute-long time interval, thus assuring repeatability conditions. Another innovative aspect of this work is that, differently from the existing literature that only use microphones in air to acquire voice samples, three devices have been employed: a calibrated sound level meter, a headworn microphone and the vocal analyzer Voice Care. The first one acts as a microphone in air that requires the subjects to remain at a fixed distance during the speech production; the second one is another microphone in air that does not impair the subject from slight movements; the third one is based on a contact microphone that senses the vocal-fold vibrations at the base of the neck.

The purposes of this work, which has been touched upon at the beginning of this paragraph, can be summarized in the following questions:

- How much SPL estimates vary within one speaker in readings acquired with the three devices? This quantity has been named as intra-speaker variability of SPL.

- How much SPL estimates vary in a group of speakers in readings acquired with the three devices? This quantity has been named as inter-speaker variability of SPL.

Further investigations are related to the influence of speech material on SPL estimates and to the effect of logging intervals on SPL variability. SPL has been separately computed on readings acquired with each device, thus allowing us to provide preliminary normative data for the assessment of results on SPL obtained in the vast majority of the study in the speech field.

3.1 Method

3.1.1 Laboratory and participants

Experiments were performed in the semi-anechoic room at the National Institute of Metrological Research (I.N.Ri.M.), where the A-weighted equivalent background noise level was 24.5 dB (33.7 dB unweighted). The mid-frequency reverberation time (from 0.5 kHz to 2 kHz) was 0.11 s. Seventeen native Italian students from Politecnico di Torino (8 males, 9 females) took part in this study (age range 19–26 years, mean age 23 years). Participants were first asked to perform an audiometric screening test using an iPad-based application titled uHear [143, 144], which provides a hearing sensitivity evaluation per frequency band (from 0.5 kHz to 6 kHz) and with a level-based rating, and they obtained results within the normal hearing level. None of them had history of speech and language disorders, based on self-report, and none of them had professional singing or speaking training.

3.1.2 Speech material

Participants were asked to read aloud two passages twice and in sequence, thus obtaining four repetitions for each subject. The speech material consisted of two standardized phonetically balanced passages (P1 and P2), which were selected being widely used for articulation drills, speech recognition testing and language studies, because they provide a broad selection of Italian-language sounds [145]. The two passages had different structures and lengths: P1 was a short tale of 300 words and

took an average reading time of about 2 minutes, while P2 was a more expressive text of 124 words and lasted about 1 minute. The texts of P1 and P2, which are reported in Appendix A, were printed on sheets and laid over a sound absorbing panel hung on a music stand, in front of the speaker's eyes, at a distance of 1 m.

The choice of asking the subjects to read was related to the need of having a continuous and various speech material that would have been the same for each participant in the experiment. Subjects were instructed not to whisper in soft voice nor to shout in loud voice, but they were advised to choose comfortable levels of loudness and pitch for a normally articulated conversational speech.

Participants performed a reading task in the same day per each person, and individual measurements were taken subsequently, in a 15-minutes time interval, thus assuring repeatability conditions.

3.1.3 Measurement set-up and procedure

The reading uttered by each subject was recorded simultaneously by means of three measurement chains, namely:

- a calibrated sound level meter (XL2 by NTi Audio), with a class 1 omnidirectional measurement microphone M2210. For the entire period of the test, each subject was asked to stand in front of the microphone, on axis, at the fixed distance of 16 cm as provided by a thin spacer. The recommended mouth-to-microphone distance for this kind of measurements is 30 cm [146, 59] and with this suggested distance, when the background noise level is lower than 25 dBA, the low-intensity voice levels can be obtained with a Signal-to-Noise Ratio (SNR) of at least 10 dB [147]. The distance was reduced to 16 cm in order to increase the SNR, since the microphone of the sound level meter is an omnidirectional one and it is not affected by the proximity effect [59], i.e. the low-frequency boost in the frequency response of a directional microphone that increases as the mouth-to-microphone distance drops;

- an omnidirectional headworn microphone Mipro MU-55HN (Chiayi, Taiwan), which was placed at a distance of about 2.5 cm from the lips' edges of the talkers, slightly to the side of the mouth, at about 20÷45 degrees horizontally, depending on the subjects' face shape. The microphone, which exhibits a

flatness of ±3 dB in the range from 40 Hz to 20 kHz, was connected to a bodypack transmitter ACT-30T, which transmits to a wireless system Mipro ACT 311. The output signal of this system was recorded with an handy recorder ZOOM H1 (Zoom Corp., Tokyo, Japan) that use a sample rate of 44.1 kHz and 16 bit of resolution. The transmitter was set without the automatic gain control, which is often responsible for SPL compression effect that may distort SPL measurements;

- a portable vocal analyzer, namely the Voice Care device (PR.O.VOICE, Turin, Italy), which was recently developed at Politecnico di Torino (see Chapter 1, paragraph 1.1.2 for details).

Figure 3.1 shows a female subject who performed the experiment and who was equipped with all the measurement devices. Before reading each passage, subjects simultaneously repeated the vowel /a/ and tapped twice the ECM with their hands in order to produce sharp peaks on the speech signals acquired by the two microphones in air and by the ECM, respectively. These peaks were considered as reference points to select signals to be analyzed in the post-processing. Among all the collected recordings (336 minutes in total), some of them were discarded in the data processing due to the failure of the preliminary calibration procedure of the Voice Care [31] and/or for incorrect execution of the experiment, (e.g. one subject moved his lips far away from the thin spacer of the sound level meter during the test). Three females performed the experiment only wearing the Voice Care. Therefore, a different number of subjects was taken into account for the three devices: 13 subjects (7 males, 6 females) were considered for the sound level meter, 14 subjects (8 males, 6 females) for the headworn microphone, and 12 subjects (7 males, 5 females) for the Voice Care device. The results were separately analyzed, since a comparison among the different devices is not the goal of the experiment.

3.1.4 SPL estimation

The stored data obtained for each participant and device was transferred to a Personal Computer and subdivided into different files, using the sharp peaks at the beginning of each reading as starting time instant for each file. This procedure was done using the software Adobe Audition (version 3.0) for the WAV audio files recorded by the sound level meter and the headworn microphone. A specific MATLAB

Fig. 3.1 From left to right: female subject while standing in front of the sound level meter XL2 by NTi Audio and wearing the headworn microphone Mipro MU-55HN and the Electret Condenser Microphone (ECM AE38) of the Voice Care device; male subject while performing the experiment in the semi-anechoic chamber of I.N.Ri.M.

(R2014b, version 8.4) script was implemented for data stored in the Voice Care device. Then, each repetition of the two passages collected per each device and subject was post-processed with specific MATLAB scripts for the estimation of speech SPL occurrences, obtaining histograms with a bin resolution of 1 dB. A speech SPL distribution was thus obtained per each reading, based on the logging interval of each device. SPLs were estimated with a logging interval of 1 s for signals acquired from the sound level meter, since it is the most common interval that is set in class 1 sound lever meters. The same logging interval was used for signals acquired from the headworn microphone. The samples acquired with the Voice Care were grouped into frames of 30 ms and a suitable *rms* voltage threshold was identified in order to distinguish voiced and unvoiced frames per each file, according to a procedure described by Carullo *et al.* [30, 31]. SPL values for voiced frames only at a fixed distance of 16 cm from the speaker's mouth was then obtained, thanks to the calibration function estimated for each subject. A calibration sine-wave file at a level of 94 dB, which was registered by coupling the sound level meter to a pressure calibrator B& K 4230, was used as a reference value in the analysis of WAV signals acquired with the sound level meter.

A different reference value was used in the analysis of data recorded by the headworn microphone and it was estimated by means of a comparative calibration procedure between the headworn microphone and the sound level meter, used as a

reference device. The characterization was performed in the anechoic chamber of Politecnico di Torino, where the measured A-weighted equivalent background noise level was 26.2 dB. Initially, the sound level meter was calibrated by coupling it to a pressure calibrator B & K 4230, which provides a nominal pressure of 1 Pa @ 1 kHz, and a calibration sine-wave file at a level of 94 dB was recorded. Then, both the sound level meter and the headworn microphone were placed at a distance of 2.5 cm from the mouth of a B & K type 4128 Head and Torso Simulator, HATS, (B & K, Nærum, Denmark), on-axis. The HATS was connected through the amplifier ALPINE MRP T222 (Alpine Electronics, Inc., Tokyo, Japan) and the audio device TASCAM US-144 (TEAC America, Inc., Montebello, CA) to a notebook PC. The software DIRAC 5 was run to generate different sound pressure levels of ICRA noise [148] in the usual range observed in professional voice users (from 55 to 72 dB @ 1 m) [48]. ICRA noise was preferred to standard signals like white or pink noise due to its speech-like spectral and temporal properties. For each sound pressure level, the output signals of the headworn microphone and the reference device were simultaneously acquired and post-processed by means of MATLAB scripts that estimated equivalent SPL, using the calibration wave file of the sound level meter as a reference. The difference between equivalent SPLs estimated from data recorded by the two devices was added to the headworn microphone levels in order to obtain calibrated values.

3.2 Analyses

SPL occurrences of each reading constituted SPL distributions that characterized each individual speech sample. The mean, mode and equivalent sound pressure levels (SPL_{mean}, SPL_{mode} and SPL_{eq}, respectively), which are the most representative descriptive parameters for the intensity of speech production, were obtained for each reading and subject. The estimation of SPL_{eq} from data acquired with the Voice Care device was performed implementing the equation proposed by Ŝvec *et al.* [18] (equation 2.1 in Chapter 2). For the estimation of SPL variability, the type A method proposed in the GUM [142] has been followed both in the experimental design and in the result processing. SPL values have been considered as random variables and SPL variability has been estimated as the experimental standard deviation of the available data.

3.2.1 Intra-speaker variability of speech SPL

With the purpose of finding the intra-speaker variability of speech SPL that occurred across readings, the experimental standard deviation of the four repeated measures for each *i-th* subject, hereafter referred as s_{i}, was calculated for SPL_{eq}, SPL_{mean} and SPL_{mode}. Then, for each device-group and SPL parameter, the average of s_{i} values ($\bar{s}$) and its 95% Confidence Interval (CI) for the mean based on a *t* critical value were calculated. This estimate has been considered as the mean descriptive parameter for intra-speaker variability in speech SPL, since it denotes, on average, the variability of vocal intensity referred to a general speaker. The *t* critical value changed depending on the number of subjects who performed the experiment with each device [142] (i.e., it was calculated as 2.18 for the sound level meter-group, 2.16 for the headworn-group and 2.20 for the Voice Care-group). A further investigation of the individual SPL variability has been performed by the estimation of the maximum differences among the four repeated measures (Δ) of SPL_{eq}, SPL_{mean} and SPL_{mode} for each subject and device.

3.2.2 Inter-speaker variability of speech SPL

Aiming to quantify the individual variability of speech SPL among speakers, also known as inter-speaker variability, the experimental standard deviation of each device-group, *s(g)*, was calculated for SPL_{eq}, SPL_{mean} and SPL_{mode}, according to the following expression:

$$s(g) = \sqrt{\frac{1}{n-1} \cdot \sum_{j=1}^{N} (r_j - \bar{r})^2} \tag{3.1}$$

where n is the number of subjects, r_{j} represents the average of the four SPL measures obtained from the four repeated readings for each subject, $\bar{s}$ is the overall mean among the r_{j} values for each device. This quantity denotes the variability of vocal intensity in a group of speakers and it has been considered as the mean descriptive parameter for inter-speaker variability in speech SPL. The standard deviation of the mean, or standard error, s_{m}, was also obtained as a ratio between *s(g)* and the root square of n, where n is the device-group sizes of the participants in the experiment. This estimate may be a reference for the investigation of changes in

speech SPL over groups of subjects or for the same group of subjects in different conditions, when a comparison between averaged measures has to be performed. It represents a significant parameter to be used as a reasonable uncertainty contribution for the mean value of the group of data, since it takes into account the group size.

3.2.3 Influence of reading material on SPL variability

Further analysis on speech SPL distributions has been conducted in order to investigate if the reading of two different passages can affect SPL variability, comparing differences in material P1 and P2 according to the voice intensity produced. For each speech SPL distribution, the following descriptive statistics were calculated: mean (SPL_{mean}), median (SPL_{median}) and mode (SPL_{mode}) as measures of location of the distribution; standard deviation (SPL_{sd}) and the interval between the maximum and the minimum value (SPL_{int}) as measures of its variance, kurtosis (SPL_{kurt}) and skewness (SPL_{skew}) for the characterization of distribution shape.

With the purpose of investigating the speech SPL distributions, the two-tailed Wilcoxon signed ranks test has been applied, that is a non-parametric test based on dependent paired samples [149]. All the descriptive statistics of the SPL distribution and SPL_{eq} were calculated for each repetition and subject involved in the study, and a two pairs were thus obtained for each subject, one related to the two readings of the first passage ($P1a - P1b$) and the other related to the two readings of the second passage (*P2a-P2b*). The average values of each SPL parameter between the two readings of each passage were also calculated for each subject (*P1m-P2m*). The Wilcoxon signed ranks test has been applied to all the paired lists of descriptive statistics for SPL distributions related to each group-device. The adopted statistical test does not require any specific assumptions on the distribution, and the null hypothesis ($H0$) states that $MD = 0$, where MD is the median of the difference between the paired sample in the two readings of each reading passage. The one-sample Kolmogorov-Smirnov test verified that data in each list did not come from a normal distribution, except for the kurtosis values of the SPL distributions (SPL_{kurt}) obtained from Voice Care, thus justifying the use of a non-parametric test for the analysis.

Table 3.1 Results on speech SPL variability obtained from the readings recorded with the calibrated sound level meter (SLM) at 16 cm from the speaker's mouth. Intra-speaker variability results: average of the individual standard deviations of SPL_{eq}, SPL_{mean} and SPL_{mode} in the four readings, $\overline{s}$, and 95% confidence interval for the mean (CI) based on a t critical value; minimum and maximum differences (Δ) of SPL_{eq}, SPL_{mean} and SPL_{mode} in the four repeated readings among subjects. Inter-speaker variability results: group mean and experimental standard deviation, $s(g)$, of SPL_{eq}, SPL_{mean} and SPL_{mode} obtained from all subjects.

Variability	SPL_{eq} (dB)		SPL_{mean} (dB)		SPL_{mode} (dB)	
Intra-speaker	$\overline{s}$ (CI)	min,max Δ	$\overline{s}$ (CI)	min,max Δ	$\overline{s}$ (CI)	min,max Δ
	0.4 (0.2-0.6)	0.2, 2.2	0.6 (0.4-0.8)	0.3, 2.6	1.0 (0.7-1.3)	1.0, 4.0
Inter-speaker	group mean	$s(g)$, s_m	group mean	$s(g)$, s_m	group mean	$s(g)$, s_m
	76.3	3.9, 1.1	74.4	3.5, 1.0	76.6	4.0, 1.1

3.2.4 Influence of logging intervals on SPL variability

Further investigations have been carried out for determining how different logging intervals can affect the speech SPL variabilities. Vocal data acquired with the sound level meter and the headworn microphone has been post-processed with a frame length of 30 ms, 250 ms and 500 ms. The same analyses described in paragraphs 3.2.1 and 3.2.2 were then performed.

3.3 Results and discussion

3.3.1 Speech SPL variability

Table 3.1 shows the results of speech SPL variability obtained from the readings that were recorded with the sound level meter at 16 cm from the speaker's mouth. SPL_{eq} shows the minimum variability within one speaker, having the minimum $\overline{s}$, that is 0.4 dB (95%-CI between 0.2-0.6 dB). Furthermore, SPL_{eq} shows the lowest range between the minimum and maximum Δ, which is equal to 2 dB, while SPL_{mode} has both the maximum intra-and inter-speaker variability, showing $\overline{s}$ equal to 1.0 dB (95%-CI between 0.7-1.3 dB) and $s(g)$ of 4.0 dB. The intra-speaker variability of SPL_{eq}, SPL_{mean} and SPL_{mode} presents values at least four times lower than those of the inter-speaker variability. The standard error s_m is equal to 1.0 dB for SPL_{mean} and to 1.1 dB for both SPL_{eq} and SPL_{mode}.

Table 3.2 The same of Table 3.1. Data refers to speech SPL obtained from the readings recorded with the headworn microphone Mipro MU-55HN at a distance of 2.5 cm from the speaker's mouth.

Variability	SPL_{eq} (dB)		SPL_{mean} (dB)		SPL_{mode} (dB)	
Intra-speaker	$\bar{s}$ (CI)	min,max Δ	$\bar{s}$ (CI)	min,max Δ	$\bar{s}$ (CI)	min,max Δ
	0.5 (0.3-0.7)	0.1, 2.3	0.6 (0.5-0.8)	0.2, 2.4	1.1 (0.7-1.5)	1.0, 5.0
Inter-speaker	group mean	$s(g)$, s_m	group mean	$s(g)$, s_m	group mean	$s(g)$, s_m
	95.1	5.0, 1.3	93.2	4.7, 1.3	95.4	5.3, 1.4

Table 3.3 The same of Table 3.1. Data refers to the readings recorded with the Voice Care, which estimates speech SPL at 16 cm from the speaker's mouth.

Variability	SPL_{eq} (dB)		SPL_{mean} (dB)		SPL_{mode} (dB)	
Intra-speaker	$\bar{s}$ (CI)	min,max Δ	$\bar{s}$ (CI)	min,max Δ	$\bar{s}$ (CI)	min,max Δ
	0.8 (0.3-1.0)	0.3, 5.2	0.6 (0.3-0.9)	0.2, 3.9	1.5 (0.8-2.2)	1.0, 9.0
Inter-speaker	group mean	$s(g)$, s_m	group mean	$s(g)$, s_m	group mean	$s(g)$, s_m
	77.9	2.8, 0.8	77.7	2.8, 0.8	79.4	3.0, 0.9

Table 3.2 shows the results of speech SPL variability obtained from the readings that were recorded with the headworn microphone Mipro MU-55HN at a distance of 2.5 cm from the speaker's mouth. SPL_{eq} shows the minimum variability within one speaker, with both the minimum $\bar{s}$ of 0.5 dB (95%-CI between 0.3-0.7 dB) and the lowest range between the minimum and maximum Δ of 2.2 dB, while SPL_{mean} shows the minimum variability among speakers, with $s(g)$ equal to 4.7 dB. SPL_{mode}e has the maximum values for both the inter- and intra-speech variability, with $\bar{s}$ equal to 1.1 dB (95%-CI between 0.7-1.3 dB) and $s(g)$ of 5.3 dB. The intra-speaker variability of SPL_{mode} and SPL_{eq} present values at least five and ten times lower than those of the inter-speaker variability, respectively. The standard error s_m is 1.3 dB for SPL_{eq} and SPL_{mean} and 1.1 dB for SPL_{mode}.

Results on SPL variability that have been obtained from readings recorded with the Voice Care, whose data refers to 16 cm from the speaker's mouth, are summarized in Table 3.3. SPL_{mean} shows the lowest intra-speaker variability, with both the minimum $\bar{s}$, that is 0.6 dB (95%-CI between 0.3÷0.9 dB), and the lowest range between the minimum and maximum Δ of 3.7 dB, while SPL_{mode} shows the highest inter-speaker variability with $s(g)$ equal to 3.0 dB. The values of intra-speaker variability are at least 3 times lower than the inter-speaker ones, except for SPL_{mode} that has the variability contributions that differ less than 1 dB. The standard error s_m is 0.8 dB for SPL_{eq} and SPL_{mean} and 0.9 dB for SPL_{mode}.

In the present study, the absolute values of the estimated SPL parameters have not been mentioned, because they are not directly included in the questions under investigation. However, Table 3.1, Table 3.2 and Table 3.3 report the group mean of each SPL parameter as complementary data for the inter-speaker variability and Figure 3.2 shows some details about the speech levels, since it summarizes for each SPL parameter and device the individual mean of the four repeated measures with the respective standard deviations (*s*) and the overall mean value with the relative experimental standard deviations, *s(g)*.

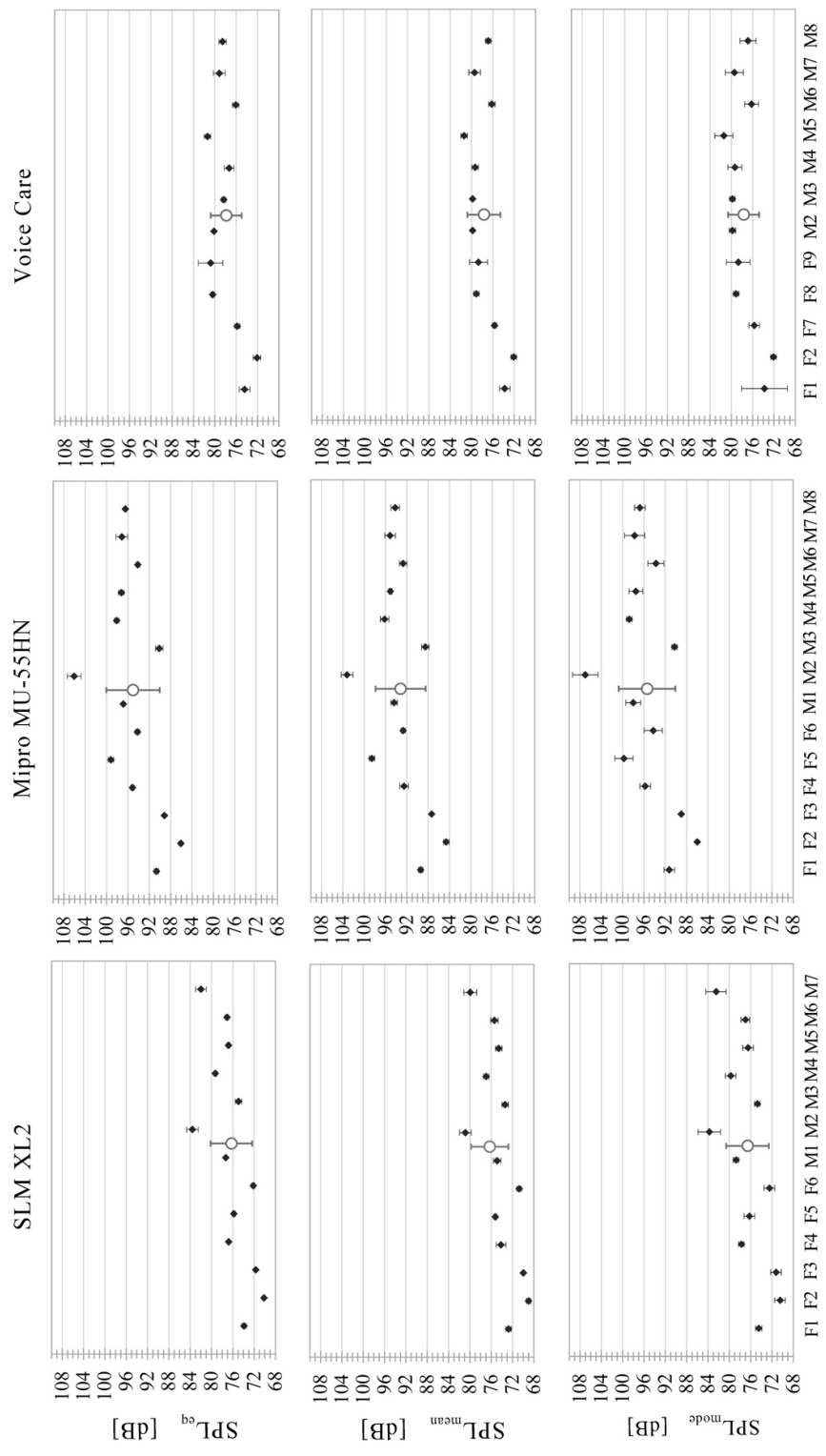

Fig. 3.2 Averaged values of SPL_{eq}, SPL_{mean} and SPL_{mode} in the two readings of the two passages for each subject, four total repetitions (diamond points); bars indicate the experimental standard deviation, s, for each subject. Overall mean value among subjects are indicated as circle points; bars indicate the experimental standard deviations, $s(g)$, of averaged values.

Due to the different computation algorithm implemented by the Voice Care to estimate SPL distributions, i.e. the vocal analyzer estimates SPLs only on voiced frames, the outcomes from the three devices have not been compared. However, some common considerations on the variability of SPL parameters in the three devices can be made. The intra-speaker variability of SPL_{eq} and SPL_{mean} results negligible for the three devices, that is within 1 dB, while it is higher for SPL_{mode}, reaching 2 dB. These outcomes are not surprising, since they reflect the type of parameter under analysis: SPL_{eq} and SPL_{mean} express average measures, while SPL_{mode} represents the most frequent observation among SPL occurrences.

The results of this study cannot be compared with most of the outcomes by other researches, because of the difference in the experimental procedure and measurement equipment. Previous works on speech SPL in readings investigated the intra-speaker variability of vocal intensity within days or within times in a day [134–136]. In the present study, instead, the evaluation has been done within successive reading tasks performed in few minutes, in order to ensure repeatability conditions. Table 3.1 shows that the inter-speaker variability of SPL_{eq} estimated from signals acquired with the sound level meter was 3.9 dB. This outcome is in agreement with the result that Corthals [137] found for the youngest group of participants in his experiment. It should be noted that, even if both the groups are made of young people, the age range of young subjects who participated to Corthals's experiment (from 7 to 17 years) did not match the one of participants who were involved in this study (from 19 to 26 years). Otherwise, the standard deviation of vocal intensity, in relative dB, that Brown et al.[134] obtained for the reading task in the young group of people, i.e. 1.9 dB, resulted definitely lower than the inter-speaker variability of SPL parameters that has been found in this work for both the sound level meter and the headworn microphone, which is shown in Tables 3.1 and 3.2, respectively.

3.3.2 Influence of reading material on SPL parameters

Table 3.4 shows the p-values obtained for each group-device, and for each paired lists of SPL parameters related to *P1a-P1b*, *P2a-P2b*, and *P1m-P2m*. None of the paired lists of quantities present significant differences across the two readings of the same passage among subjects (p-values > 0.05), with the exception of SPL_{kurt} for the readings of the second passage acquired with the Voice Care. A main result of this analysis is that, generally, each subject in each device-group repeated the reading of

Table 3.4 *P*-values of the two-tailed Wilcoxon signed ranks test of the paired lists of descriptive statistics for the sound pressure level (SPL) distributions and SPL_{eq}, related to the repetitions of the first passage (P1a, P1b) and of the second passage (P2a, P2b). *P*-values refers also to pooled data from the two readings (P1m, P2m).Values lower than a significance level of 0.05 are in bold and indicate the rejection of the null hypothesis *H0*: $MD = 0$, where *MD* is the median of the population of the differences between the paired sample data.

Device	Passage	SPL_{mean} (dB)	SPL_{sd} (dB)	SPL_{median} (dB)	SPL_{mode} (dB)	SPL_{kurt} (dB)	SPL_{skew} (dB)	SPL_{int} (dB)	SPL_{eq} (dB)
SLM	*P1a-P1b*	0.556	0.464	0.125	0.828	0.588	0.984	0.840	0.151
	P2a-P2b	0.576	0.852	0.625	0.305	0.305	0.210	0.721	0.490
	P1m-P2m	0.924	**0.040**	0.250	0.117	**0.040**	0.124	**0.034**	0.138
Mipro	*P1a-P1b*	0.571	0.424	0.188	0.766	0.658	0.886	0.307	0.140
	P2a-P2b	0.690	0.572	1.000	0.090	0.391	0.199	0.764	0.419
	P1m-P2m	0.653	**0.010**	**0.002**	**0.035**	**0.016**	0.092	**0.035**	**0.009**
Voice Care	*P1a-P1b*	0.938	0.231	1.000	0.654	0.727	1.000	0.510	0.787
	P2a-P2b	0.574	0.924	1.000	0.941	**0.312**	0.176	0.488	0.639
	P1m-P2m	0.310	0.197	0.766	0.199	0.360	1.000	0.214	0.916

the same passage with similar speech levels. On the other hand, from the analysis of the paired lists of *P1m* and *P2m* significant differences have been found. SPL_{sd}, SPL_{kurt} and SPL_{int} significantly change in readings of the two passages acquired with the sound level meter. In the case of the headworn microphone, SPL parameters corresponding to P1m that result significantly different from SPL parameters obtained by P2m are SPL_{sd}, SPL_{median}, SPL_{mode}, SPL_{kurt}, SPL_{int} and SPL_{eq}. None of the SPL parameters significantly changes for the Voice Care. These outcomes reveal that subjects recorded with the sound level meter and the headworn microphone tended to read the two passages with different sound speech levels. Therefore, a negligible intra-speaker variability would be expected in the repetition of the same passage, but non-negligible intra-speaker variability could be expected between the readings of the two different passages. These findings validates our choice about the experiment, since two readings of two different passages may guarantee a sufficiently diversified speech material.

The *p*-values of the two-tailed Wilcoxon signed rank test of the paired lists of descriptive statistics for SPL distributions and SPL_{eq}, related to the repetitions of the first and second passage, reveal additional aspects of speech SPL variability: people tends to read the same passage without variations in SPL, i.e. p-value > 0.05 for *P1a-P1b* and *P2a-P2b* pairs, while they have a tendency to read different speech materials with altered SPL, i.e. some p-value < 0.05 occurred for *P1m-P2m*. However, the

reading order of P1 and P2 was not counterbalanced in the subjects, so that a time recording effect can happen, that is the two readings of P2 always followed the two readings of P1 and P1m and P2m had different SPL due to an effect of either speaker fatigue or speaker habituation to the recording environment. An additional aspect is that P2 is a more expressive passage than P1. Therefore, the one-left-tailed Wilcoxon signed rank test has been performed to the SPL paired lists. SPL_{sd} for the sound level meter, SPL_{sd}, SPL_{median}, SPL_{mode} and SPL_{eq} for the headworn microphone have p-value < 0.05. In other words, for these two device-groups, values of such SPL parameters in the first passage resulted significantly lower than those in the second passage, thus having the presence of a certain time recording effect or the more expressive nature of the second passage as possible reasons. None of SPL parameters has p-value < 0.05 for the Voice Care. The results obtained for the Voice Care from both the two-tailed and the one-tailed Wilcoxon signed ranks test give an indication that pauses, i.e. *unvoiced frames* that are discarded in the process algorithm, are relevant in the distribution of SPL.

3.3.3 Influence of logging intervals on SPL variability

Table 3.5 shows results on speech SPL variability obtained by post-processing the readings acquired with the sound level meter and the headworn microphone with different logging intervals. The intra-speaker variability, $\overline{s}$, for SPL_{eq} keeps constant by post-processing the reading samples with logging intervals equal to 1000 ms, 750 ms, 500 ms, 250 ms and 30 ms, both for the sound level meter and the headworn microphone. For both the microphones, SPL_{mean} has a deviation of 0.1 dB among $\overline{s}$ values, while SPL_{mode} shows an upward trend of $\overline{s}$ when logging intervals decrease. An extreme result of 6.4 dB can be easily noticed in the $\overline{s}$ values obtained from data acquired with the headworn microphone and post-processed with a logging interval of 30 ms. It is due to the conjunction of two phenomena, that are the use of 30 ms-frame length and the internal noise of the measurement chain of the headworn microphone. Such a frame length is short enough to obtain several SPL occurrences of the unvoiced frames, which could have SPL values similar to the background noise in the semi-anechoic chamber. This assumption has been confirmed, since a silent period of 10 s was recorded with the headworn microphone and the equivalent level was equal to 46 dB, which actually is the internal noise of the headworn microphone.

Table 3.5 Results of speech SPL variability obtained by post-processing the reading voice signals of readings with different logging intervals. Speech samples are recorded with the calibrated sound level meter (SLM) XL2 at 16 cm from the speaker's mouth and with the headworn microphone Mipro MU-55HN at a distance of 2.5 cm from the speaker's mouth. Intra-speaker variability results: average of the individual standard deviations, $\bar{s}$, of SPL_{eq}, and the mean SPL_{mean} and mode SPL_{mode} in the four readings. Inter-speaker variability results: experimental standard deviation, *s(g)*, of SPL_{eq}, SPL_{mean} and SPL_{mode} obtained from all subjects.

SPL parameter (dB)	Logging interval (ms)	SLM		Mipro	
		$\bar{s}$	*s(g)*	$\bar{s}$	*s(g)*
SPL_{eq}	1000	0.4	3.9	0.5	5.0
	750	0.4	3.9	0.5	5.0
	500	0.4	3.9	0.5	5.0
	250	0.4	3.9	0.5	5.0
	30	0.4	3.9	0.5	5.1
SPL_{mean}	1000	0.6	3.5	0.7	4.7
	750	0.6	3.4	0.7	4.5
	500	0.7	3.3	0.7	4.5
	250	0.7	3.1	0.8	4.3
	30	0.7	2.7	0.8	3.9
SPL_{mode}	1000	0.9	4.0	1.1	5.3
	750	1.0	4.1	1.1	4.9
	500	1.1	4.1	1.3	4.9
	250	1.2	4.1	1.4	5.0
	30	1.3	3.9	**6.4**	**13.9**

Figure 3.3 shows two distributions of SPL occurrences, which both exhibit a bimodal shape, obtained using a 30 ms logging-interval of a reading that was simultaneously acquired with the sound level meter and the headworn microphone. The SPL distribution that refers to the sound level meter has the lowest peak-level equal to 34 dB, while the SPL distribution of the headworn microphone has the lowest peak-level equal to 48 dB. As highlighted by Hodgson et al.[124], the lowest-peak level of a long-term speech corresponds to the background noise that occurs during the voice monitoring. The SPL distribution obtained from the sound level meter reflects this finding, since a correspondence between the lowest peak-level (34 dB) and the background noise that was measured in the semi-anechoic chamber (33.7 dB, as reported in 3.1) has been found. For the headworn microphone, a difference of 2 dB between the internal noise and the lowest peak-level occurs. However, it seems that occurrences of both low SPLs and internal noise have been accumulated at 48 dB, determining the highest peak-level in correspondence of that value. This phenomenon results in 7 out of 56 SPL distributions with the highest occurrence near to the internal noise level, thus achieving the extreme $\bar{s}$ value of 6.4 dB.

Table 3.5 also shows results on the inter-speaker variability of SPL parameters in the two microphones. Despite the varying of logging intervals, *s(g)* remains the same for SPL_{eq} with a deviation of 0.1 dB for the headworn microphones, while it shows a downward trend when logging intervals decrease both in SPL_{mean} and SPL_{mode}, with the exception of the extreme value of *s(g)* that corresponds to the 30 ms-logging interval in SPL_{mode}, that is 13.9 dB. This anomalous *s(g)* behavior can be attributed to the same phenomenon that has been explained above. It should be noted that SPL distributions from reading samples of 1 minute using logging intervals of 1 s have only 60 points, thus their descriptive statistics could have some random variation linked to the low number of data points, especially for SPL_{mode}. With shorter frame durations, instead, a greater number of points is included in the histogram of the measured data, thus allowing an underlying theoretical random distribution to be estimated, so that SPL_{mode} becomes a good estimator of the peak of the distribution. Despite this, the intra-speaker variability of SPL_{mode} increases as logging intervals decreases, while a not clear trend has been identified for the inter-speaker variability of SPL_{mode}. SPL_{mean} reveals slight changes in the intra-speaker variability and a decrease of the inter-speaker variability as logging intervals become shorter. As highlighted by Ŝvec *et al.*[18], there is the evidence that different SPL_{mean} values can be obtained for the same voice signal when different logging intervals are used in the

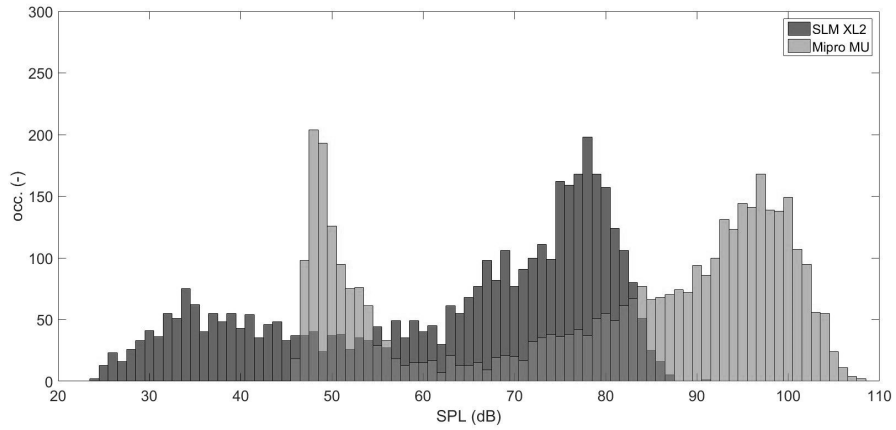

Fig. 3.3 Two distributions of SPL occurrences obtained from the analysis of a reading that was simultaneously acquired with both the SLM XLS (dark grey) and the headworn microphone Mipro MU-55HN (light grey). The logging interval used in the post-processing was 30 ms.

analysis, so that modifications in SPL_{mean} variabilities can be expected. Eventually, both the intra- and the inter-speaker variability of SPL_{eq} keep quite constant as logging intervals change, according to its definition of time-weighted average of SPLs.

3.4 Instruction of use

The results reported in the present study may be affected by the lower reproducibility due to the relative position between the subject and the devices during the experiment. For the sound level meter, subjects could have slightly moved their lips from the thin spacer during the readings. The arch of the headworn microphone is crucial for two main reasons: it could have slightly changed the distance from the lips and the microphone during the experiment because of its thinness and it has a fixed length that caused a different horizontal angle from the mouth, depending on the subjects' shape of face. Therefore, the microphone could be placed in the airflow area for some subjects, thus acquiring unwanted artefacts despite the use of the windscreen. Further precautions are needed in future research. In addition, the estimation of SPL from the wearable vocal analyzer is affected by the sensitivity of the ECM with respect to body activity, the so-called tissue-borne effects, which could occur

during voice monitoring. It provides an additional contribution of uncertainty in the estimation of speech SPL [31].

The outcomes in the present study are preliminary, mainly because of the limited number of subjects who took part in the experiment. Further researches should involve more subjects and it could be useful to ask the speakers to repeat more than four readings, in order to obtain more reliable values from individual standard deviations.

It is also important to consider the application of these preliminary types of normative data. The results of the intra-speaker variability may be particularly useful in studies that investigate individual differences in speech SPL, which can be measured in two different periods or conditions. The outcomes of the inter-speaker variability may be a reference for the investigation of changes in speech SPL over groups of subjects. When a comparison between averaged measures among groups of subjects have to be performed, researchers may refer to *s(g)* values given in this study and calculate the standard deviation of the mean (s_{m}), or standard error, which can be obtained as a ratio between *s(g)* and the root square of n, where n is the group size of the participants in the experiment. It is important to underline that the use of values given in this paper is limited to situations in which the equipment and experimental set-up are the same as those in the present study. Researchers often make comparisons between different situations, e.g. states of health and room acoustic conditions, and evaluate SPL trends in a subject or among groups of subjects in long-term monitorings. As a general rule, when differences are greater than $\overline{s}$ and *s(g)* (or s_{m} for averaged measures), it can be assumed that the new aspect that changes the previous situation has a significant influence on the intensity in speech production in a single subject or in groups of subjects, respectively. Figures 3.4 and 3.5 illustrate the two general situations with a speaker and a group of subjects.

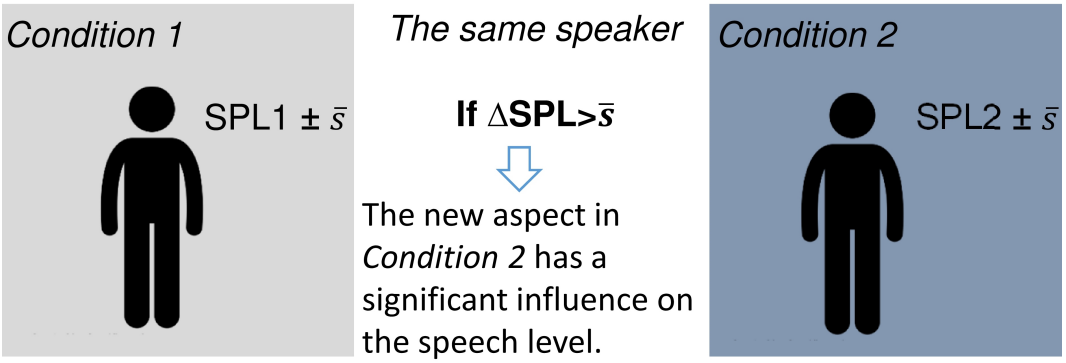

Fig. 3.4 Instruction of use for the intra-speaker variability of SPL parameters when the same subject speaks in two different conditions.

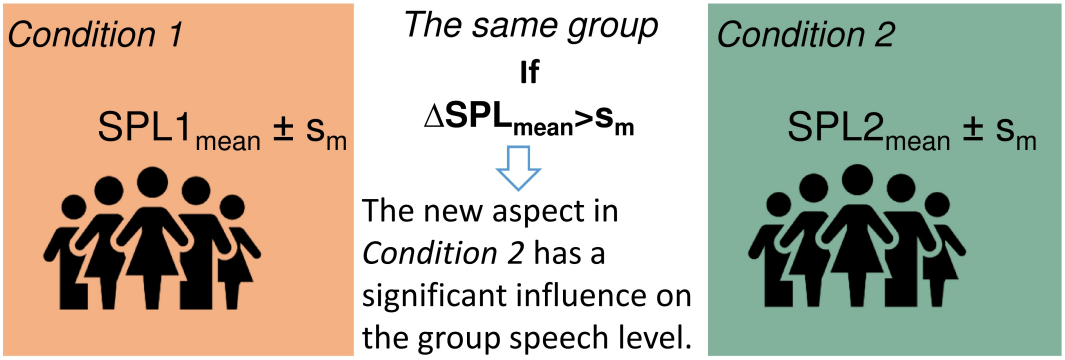

Fig. 3.5 Instruction of use for the inter-speaker variability of SPL parameters when the same group of subjects speak in two different conditions.

Chapter 4

In laboratory investigations on speech sound pressure level

This chapter partially reports material from:

1. A. Astolfi, A. Castellana, A. Carullo and G.E. Puglisi, *Measurement uncertainty of speech level and speech level difference for a contact-sensor-based device and a headworn microphone*, The Journal of the Acoustical Society of America-Express Letter, *submitted*.

2. A. Astolfi, A. Castellana, G.E. Puglisi, A. Carullo, U. Fugiglando, *Investigation on the effects of very low and excessive reverberation in speech levels*, The Journal of the Acoustical Society of America, *to be submitted*.

The contents of this chapter are related to in-laboratory experiments that further investigate speech level measures provided by different types of microphone.

In light on the considerations of paragraph 1.1.4, the first part of this chapter provides guidelines for estimating each uncertainty contribution that affects speech SPL measures obtained with a contact-sensor based device as well as with a headworn microphone. In particular, the uncertainty contributions for absolute measures of instantaneous speech sound pressure level, SPL_{i}, and for absolute measures and differences of equivalent, SPL_{eq}, mean, SPL_{mean}, and mode, SPL_{mode}, speech sound pressure level are calculated for the two devices.

Based on these findings, the second part of the chapter investigates changes in the voice intensity while speaking in rooms with very low and very high reverberation

time. In particular, the increase in the sound pressure level parameters and in the sound power level in a semi-anechoic room compared to a reverberant room has been assessed with the contact-based microphone vocal analyzer and the headworn microphone, for free speech and a map description speech tasks.

4.1 Uncertainty estimation of speech level measures

4.1.1 Method

Two devices have been used in this work: the Voice Care device, which is a contact-sensor based device as detailed in paragraph 1.1.2, and the omni-directional headworn microphone Mipro MU-55HN as described in paragraph 3.1.3. The Mipro MU-55HN is usually placed at about 2.5 cm from the lips' edge of a talker, slightly to the side of the mouth, at about 20°-40° horizontally depending on a subject's head size. The microphone is connected to a bodypack transmitter ACT-30T that transmits to a wireless microphone system Mipro ACT 311 and Wav signals are stored on a handy recorder ZOOM H1 (Zoom Corp., Tokyo, Japan) in 16 bits/44.1 kHz format. A logging interval of 1 s was used in the post processing with an *ad hoc* MATLAB (R2014b, version 8.4) script for the estimation of the SPL_{i} values. This interval was chosen since it is the most common in measurements with such devices. Moreover, in the case of Mipro MU-55HN, a logging interval lower than 1 s and comparable with the intersyllabic pause of 30 ms to 60 ms in Italian language [32, 33], would bring to a bimodal distribution for SPL_{i}, where the peak at lower levels due to background noise could overcome the peak at higher levels, due to speech (see Chapter 4 for details) [60].

The Mipro MU-55HN was calibrated against a reference NTi XL2 class 1 sound level meter (SLM) equipped with an ominidirectional M2210 microphone (NTi Audio, Schaan, Liechhtenstein). The Guide to the expression of Uncertainty in Measurement [142] is the main reference used to deal with the uncertainty of experimental data. According to this document, the uncertainty for SPL_{i} values detected by the Voice Care device and by the headworn microphone Mipro MU-55HN can be obtained combining the different uncertainty contributions that affect the measurement procedure. In this study, three main uncertainty contributions have been considered, that are, instrumental uncertainty, method reproducibility and method repeatability.

The instrumental uncertainty is mainly due to contributions related to the calibration of the device and to its verification against a standard microphone. The method reproducibility uncertainty is related to the closeness of the agreement between the results of measurement sessions of the same measurand carried out under changed conditions (i.e. experimental set-up and influence quantities). Method repeatability uncertainty considers repeated measurements in the same nominal conditions. Under the assumption of uncorrelated uncertainty contributions, the estimation of SPL_{i} standard uncertainty can be obtained as:

$$u(SPL_{\mathrm{i}}) = \sqrt{u(SPL_{\mathrm{i,inst}})^2 + u(SPL_{\mathrm{i,repr}})^2} \tag{4.1}$$

where $u(SPL_{\mathrm{i,inst}})$ is the instrumental contribution and $u(SPL_{\mathrm{i,repr}})$ is the method reproducibility contribution. In equation 4.1 the repeatability contribution is not present, since it is included in the reproducibility that was estimated performing multiple readings under changed conditions. The expanded uncertainty, $U(SPL_{\mathrm{i}})$, is then calculated from the standard uncertainty and the coverage factor k, assumed equal to 2, according to the following formula:

$$U(SPL_{\mathrm{i}}) = k \cdot u(SPL_{\mathrm{i}}) \tag{4.2}$$

Long-term speech monitorings are usually characterized in terms of SPL parameters, SPL_{par}, namely equivalent, SPL_{eq}, mean, SPL_{mean}, and mode, SPL_{mode}, sound pressure level. Since the instrumental contribution of SPL_{par} is due to systematic effects, it is considered equal to the one estimated for SPL_{i}. The standard uncertainty of SPL_{par} absolute measures due to the random contributions of method reproducibility and repeatability is evaluated using the following relationship based on SPL_{i} [142]:

$$U(SPL_{\mathrm{par}}) = \sqrt{\sum_{i=1}^{N} \left(\frac{\partial SPL_{\mathrm{par}}}{\partial SPL_{\mathrm{i}}} \right)^2 \cdot u(SPL_{\mathrm{i}})^2} \tag{4.3}$$

The parameter SPL_{mean} is estimated according to the following expression [18]:

$$SPL_{\mathrm{m}} = \frac{\sum_{i=1}^{N} n \cdot SPL_{\mathrm{i}}}{\sum_{i=1}^{N} n_i} \tag{4.4}$$

where N is the total number of frames in the analyzed speech and n is equal to 0 for the unvoiced frames and 1 for the voiced frames, in the case of Voice Care, while it is equal to 1 for all the frames in the case of the headworn microphone. According to equation 4.3, the uncertainty contributions due to method reproducibility and method repeatability for absolute measures of SPL_{mean}, $u(SPL_{\text{m}})$, are estimated as follows:

$$u(SPL_{\text{m}}) = \frac{u(SPL_{\text{i}})}{\sqrt{\sum_{i=1}^{N} n_i}} \tag{4.5}$$

The parameter SPL_{eq} is calculated using the following equation according to Švec *et al.* (2005):

$$SPL_{\text{eq}} = 10log\left(\frac{1}{N}\sum_{i=1}^{N} n_i \cdot 10^{\frac{SPL_{\text{i}}}{10}}\right) \tag{4.6}$$

and its uncertainty contributions due to method reproducibility and repeatability for absolute measures of SPL_{eq}, $u(SPL_{\text{eq}})$, is obtained as follows, according to equation 4.3:

$$u(SPL_{\text{eq}}) = \frac{u(SPL_{\text{i}})}{\sqrt{N}} \tag{4.7}$$

Since SPL_{mode} represents the most occurring value in the SPL_{i} distribution, the standard uncertainty for SPL_{mode} has been considered equal to the uncertainty of SPL_{i}. Another cause of uncertainty of speech SPL measurements is the reproducibility due to the variability of the human speech, i.e. the source reproducibility, as evaluated in Chapter 3 [60]). This contribution is meaningful only in the case of SPL_{par}, since in the case of SPL_{i} it only expresses the spread of speech SPL_{i} distribution. Eventually, the standard uncertainty of SPL_{par} absolute measures is obtained from the combination of the instrumental contribution, $u(SPL_{\text{i,inst}})$, the method reproducibility contribution, $u(SPL_{\text{par,repr}})$, and the source reproducibility contribution, $u(SPL_{\text{par,reps}})$, as follows:

$$u(SPL_{\text{par}}) = \sqrt{u(SPL_{\text{i,inst}})^2 + u(SPL_{\text{par,repr}})^2 + u(SPL_{\text{par,reps}})^2} \tag{4.8}$$

The uncertainty due to the method repeatability is included in the method reproducibility. In the case of differences between two SPL_{par} measures, the standard uncertainty only includes method repeatability, $u(SPL_{\text{par,repe}})$, and source reproducibility contribution, $u(SPL_{\text{par,reps}})$, as follows:

$$u(SPL_{\text{par}}) = \sqrt{u(SPL_{\text{par,repe}})^2 + u(SPL_{\text{par,reps}})^2} \tag{4.9}$$

The instrumental uncertainty contribution is assumed to be negligible, thus it has not been included in formula 4.9, provided that the measurements of the two quantities are performed with the same instrument in a short time interval and in similar conditions for the influence quantities. The uncertainty contribution due to method reproducibility is generally evaluated through repeated sessions that consider the replacement of the measurement chain. When the replacement of the device is not needed for repeated sessions, as for the assessment of differences between quantities, the uncertainty contribution due to method reproducibility is confined to the uncertainty contribution due to repeatability only. The contributions $u(SPL_{\text{par,repr}})$ and $u(SPL_{\text{par,repe}})$ in formulas 4.8 and 4.9 are calculated according to the expressions 4.5 and 4.7) for SPL_{mean} and SPL_{eq}, respectively. The expanded uncertainty for speech sound pressure level parameters, $U(SPL_{\text{par}})$, is then calculated from the standard uncertainty, $u(SPL_{\text{par}})$, according to equation 4.2.

Instrumental uncertainty

Voice Care In the case of the Voice Care device, the SPL_{i} instrumental uncertainty, $u(SPL_{\text{inst}})$, was estimated combining two contributions in the same way as in equation 4.1. First, the standard uncertainty due to the calibration of the reference microphone Behringer ECM8000 with the sound level calibrator, $u(SPL_{\text{inst,ref}})$, obtained as described by equation (6) in Carullo *et al.* [31] was considered. Second, the uncertainty related to the estimation of the calibration function of the Voice Care device, $u(SPL_{\text{inst,cal}})$, which is also called "model error" was taken into account.

The model error was obtained using the phonatory system simulator described in Casassa *et al.* [150] as a source: the reference microphone senses the voice signal at the output of the 3D-printed hollow resonator that simulates the vocal tract, while the contact sensor is attached to the phantom material that mimics skin tissues and muscles at the jugular notch. The simulator was driven by an EGG signal recorded

during in-vivo acquisition of a vowel /a/ at increasing intensity, thus replicating the calibration procedure defined for the device. Three calibration sessions, including 5 repetitions each, were performed in a quiet dead room with background noise lower than 35 dB(A) L_{Aeq}. For each session the measurement set up was repositioned. The model error was calculated as the maximum value, over the 15 calibrations, of the rms of the difference between $SPL_{\mathrm{i,ref}}$ and SPL_{i}, estimated through the fitted calibration function.

Headworn microphone For the headworn microphone Mipro MU-55HN, the SPL_{i} instrumental uncertainty has been estimated from the combination, in the same way as in equation 4.1, of the standard uncertainty due to the calibration of the reference class 1 SLM, $u(SPL_{\mathrm{inst,ref}})$, with the error between the SPL_{i} measures provided by the headworn microphone and the reference microphone, $u(SPL_{\mathrm{inst,delta}})$. The $u(SPL_{\mathrm{inst,ref}})$ was assumed from the calibration certificate provided by the manufacturer. In order to estimate $u(SPL_{\mathrm{inst,delta}})$, the linearity and the absence of compression effects of the headworn-microphone chain in the wide SPL_{i} range (85÷116) dB @ 2.5 cm were previously checked, after the automatic gain control were excluded. Thereafter, both the microphones were positioned in front of a B & K 4128 Head And Torso Simulator, HATS, which emitted samples of ICRA noise [148], 20 s long, in the range (85÷102) dB @ 2.5 cm. This range corresponds to (53÷70) dB @ 1 m in free field, i.e. from "normal" to "loud" vocal effort according to the ANSI S3.5 [61]. The contribution due to the measurement error was estimated over 4 differences, equally distributed in the selected range, between $SPL_{\mathrm{eq,ICRA}}$ provided by the headworn microphone and the SLM. SPL_{eq} has been used in place of the instantaneous SPL value since it is the best indicator to be used in the case of stationary signals, as ICRA. An offset value, which is the average of the 4 $SPL_{\mathrm{eq,ICRA}}$ differences, was used as a correction factor for each $SPL_{\mathrm{eq,ICRA}}$ provided by the headworn microphone. A uniform probability density function was then used to characterize each difference corrected with the estimated offset, thus allowing its standard deviation σ to be determined as $\sigma = \frac{\Delta SPL_{\mathrm{eq,ICRA}}}{\sqrt{3}}$, where $\Delta SPL_{\mathrm{eq,ICRA}}$ is the absolute value of each corrected difference. The maximum standard deviation σ over the 4 $SPL_{\mathrm{eq,ICRA}}$ differences corresponds to $u(SPL_{\mathrm{inst,delta}})$.

Method reproducibility

Voice Care In the case of the Voice Care device, the SPL_{i} uncertainty contribution due to the method reproducibility, $u(SPL_{\mathrm{i,repr}})$, was estimated as the maximum spread among all the 15 calibration functions identified during the 3 sessions performed with the phonatory system simulator. A peak to peak value has been obtained among the 15 functions, $\Delta\ SPL_{\mathrm{i,peak}}$, thus allowing the corresponding standard uncertainty to be estimated assuming a uniform probability distribution:

$$u(SPL_{\mathrm{rep}}) = \frac{\Delta SPL_{\mathrm{i,peak}}}{2 \cdot \sqrt{3}} \tag{4.10}$$

The uncertainty contribution due to the calibration repeatability, $u(SPL_{\mathrm{i,repe}})$, was instead estimated applying equation 4.10 on the highest $\Delta SPL_{\mathrm{i,peak}}$ value obtained among the sessions, which included 5 calibrations each.

Headworn microphone For the headworn microphone Mipro MU-55HN, the SPL_{i} uncertainty contribution due to the method reproducibility, $u(SPL_{\mathrm{i,repr}})$, was estimated in the semi-anechoic room of the National Institute of Metrological Research (I.N.Ri.M.) in Turin, using the HATS as speech source. The source emitted ICRA noise at a fixed gain, with the headworn microphone at about 2.5 cm from the HATS' mouth, slightly to the side. Three different sessions, including 4 repetitions each, were performed, where the measurement set up was replaced. The uncertainty contribution of reproducibility was evaluated as standard deviation of $SPL_{\mathrm{eq,ICRA}}$ over the 12 repetitions. Such evaluation implicitly includes the negligible uncertainty contribution of repeatability, which was calculated as standard deviation of the 4 repetitions of each session.

Source reproducibility

The results obtained in Chapter 3 as intra-speaker variability of speech SPL_{par}, i.e. the uncertainty contribution related to the source reproducibility of a single speaker, using both Voice Care and the headworn microphone have been considered.

4.1.2 Results

Table 4.1 shows the standard uncertainties and the expanded uncertainty for SPL_{i} values detected by Voice Care and Mipro MU-55HN. Table 4.2 shows the standard and expanded uncertainties for absolute values and differences of SPL_{eq}, SPL_{mean} and SPL_{mode} measures, detected by the two devices.

Table 4.1 Instrumental $u(SPL_{\mathrm{i,inst}})$ and method reproducibility $u(SPL_{\mathrm{i,repr}})$ standard uncertainty contributions, and expanded uncertainty $U(SPL_{\mathrm{i}})$, for instantaneous sound pressure level, SPL_{i} (dB), detected by the Voice Care voice monitoring device and the headworn microphone Mipro MU-55HN.

	Voice Care	**Headworn microphone**
	SPL_{i}	
$u(SPL_{\mathrm{i,inst}})$	1.2	0.7
$u(SPL_{\mathrm{i,repr}})$	0.5	0.1
$U(SPL_{\mathrm{i}})$	**2.6**	**1.4**

Table 4.2 Instrumental $u(SPL_{\mathrm{i,inst}})$, method repeatability $u(SPL_{\mathrm{par,repe}})$, method reproducibility $u(SPL_{\mathrm{par,repr}})$ and source reproducibility $u(SPL_{\mathrm{par,reps}})$ standard uncertainty contributions, and expanded uncertainty $U(SPL_{\mathrm{par}})$, for absolute measures and differences between measures of equivalent, mean and mode sound pressure level (dB), detected by the Voice Care voice monitoring device and the headworn microphone Mipro MU-55HN.

	Voice Care			**Headworn microphone**		
	SPL_{eq}	SPL_{mean}	SPL_{mode}	SPL_{eq}	SPL_{mean}	SPL_{mode}
	Absolute measures					
$u(SPL_{\mathrm{i,inst}})$	1.2	1.2	1.2	0.7	0.7	0.7
$u(SPL_{\mathrm{par,repr}})$	0.01	0.01	0.5	0.01	0.01	0.1
$u(SPL_{\mathrm{par,reps}})$	0.8	0.6	1.5	0.5	0.6	1.1
$U(SPL_{\mathrm{i}})$	**2.9**	**2.7**	**4.0**	**1.7**	**1.8**	**2.6**
	Differences between measures					
$u(SPL_{\mathrm{par,repe}})$	0.01	0.01	0.3	0.01	0.01	0.01
$u(SPL_{\mathrm{par,reps}})$	0.8	0.6	1.5	0.5	0.6	1.1
$U(SPL_{\mathrm{par}})$	**1.6**	**1.2**	**3.1**	**1.0**	**1.2**	**2.1**

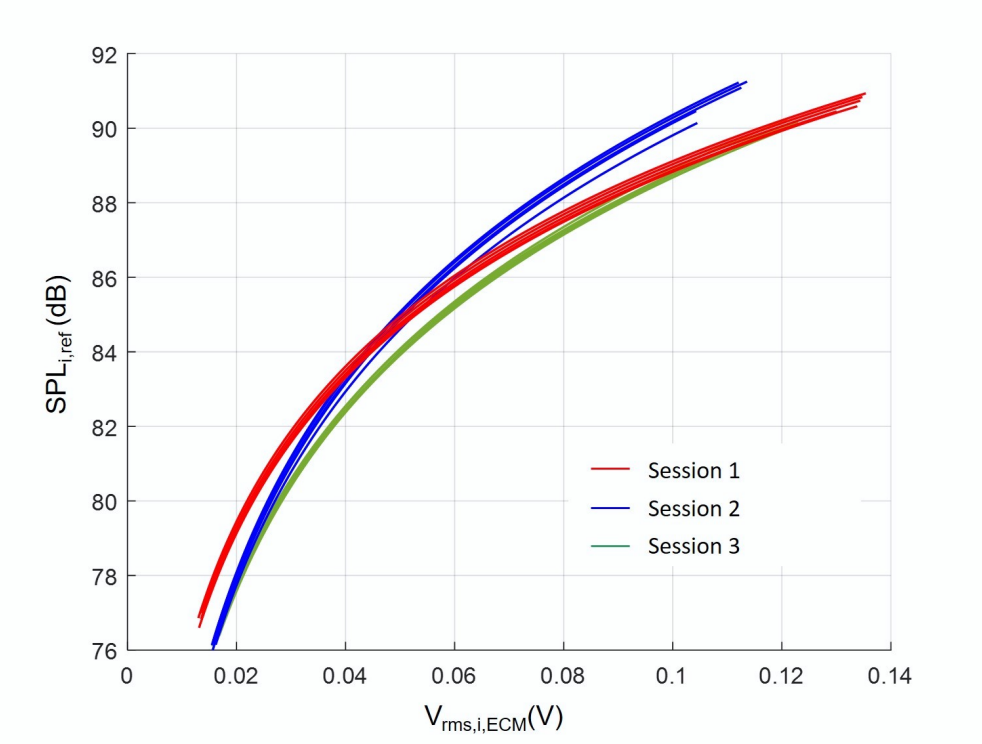

Fig. 4.1 Calibration functions performed in three calibration sessions with the Voice Care device, including 5 repetitions each. For each session the measurement set up was repositioned. $SPL_{i,ref}$ refers to 13 cm from the phonatory system simulator mouth.

Instrumental uncertainty

For the Voice Care device, the standard instrumental uncertainty was evaluated in the SPL_i range (76÷91) dB @ 13 cm, i.e. (90÷105) dB @ 2.5 cm and (58÷73) dB @ 1 m assuming a free field propagation. The uncertainty contribution due to the calibration of the reference microphone, $u(SPL_{inst,ref})$, in the $SPL_{i,ref}$ range (76÷91) dB @ 13 cm resulted from 0.6 dB to 0.3 dB, respectively. The model error, $u(SPL_{inst,cal})$, was obtained from data in Figure 4.1, which shows the calibration functions performed in the three calibration sessions, including 5 repetitions each. The maximum value of the rms of the difference between the $SPL_{i,ref}$ measured by the reference microphone and the SPL_i estimated by the calibration function was equal to 1 dB. Eventually, the combined SPL_i instrumental uncertainty, $SPL_{i,inst}$, due to the Voice Care device instrumental issues, resulted equal to 1.2 dB, when the maximum values of the related contributions have been considered.

In the case of Mipro MU-55HN, a standard uncertainty due to the calibration of the reference sound level meter, $u(SPL_{\text{inst,ref}})$, equal to 0.55 dB, was assumed from the calibration certificate. About the error between the measures provided by the headworn microphone and the reference microphone, $u(SPL_{\text{inst,delta}})$, a maximum standard deviation of 0.50 dB in the investigated range (85÷102) dB @ 2.5 cm was found considering 4 differences between $SPL_{\text{eq,ICRA}}$ provided by the headworn microphone and the reference microphone. The combination of the two uncertainty contributions provides a value of 0.74 dB as standard instrumental uncertainty of the headworn microphone, $u(SPL_{\text{i,inst}})$. This procedure does not take possible drift effects into account, which are considered negligible since the headworn microphone was used within a short time interval from the calibration. When SPL_{i} values are considered, the greatest contribution to the expanded uncertainty is the instrumental one, for both Voice Care and the headworn microphone. This is confirmed in the case of absolute measure of SPL_{par} for Voice Care, while for Mipro MU-55HN the instrumental and the source reproducibility contributions are comparable. The instrumental uncertainty contribution in the case of Voice Care is double with respect to the headworn microphone, and this is due to the uncertainty in the identification of the calibration function that also affects the higher method reproducibility contribution for SPL_{i}.

As far as Voice Care is concerned, the contribution due to the identification of the calibration function takes into account the model error (fitting contribution) that is equal to 1 dB. This uncertainty contribution is lower than the values obtained by Carullo *et al.* [31], considering a male and a female human sources uttering the vowel /a/, which are in the range 1.4÷2.7 dB. Such higher values are due to the inclusion of human sources in the experiment, thus not allowing the effect of the source reproducibility contribution to be separated from the model error. For the estimation of the model error by means of the phonatory system simulator, only the vowel /a/ was used in this study. Even though vowel /a/ is the speech material more referred in literature for this type of measurements [27, 43, 26, 19], further research is needed to collect model errors in case of other vowels or continuous speech. As example, Carullo *et al.* [31] found slightly lower values in the case of a short sentence.

Method reproducibility and repeatability

In the case of Voice Care, the SPL_{i} uncertainty contribution due to method reproducibility, $u(SPL_{\mathrm{i,repr}})$, and repeatability, $u(SPL_{\mathrm{i,repe}})$, were equal to 0.5 dB and 0.3 dB, respectively. The method repeatability and reproducibility uncertainty contributions for SPL_{eq} and SPL_{mean}, $u(SPL_{\mathrm{par,repe}})$ and $u(SPL_{\mathrm{par,repr}})$ respectively, were estimated assuming a typical occupational speech monitoring time of 2 minutes and a phonation time percentage of 30% [60, 48]. In the case of Voice Care, 1200 voiced frames (sum of voiced frames) were obtained over a total number of 4000 frames (N). According to equation 4.5, the SPL_{mean} method repeatability uncertainty resulted lower than 0.01 dB, while the method reproducibility uncertainty was of 0.01 dB. According to equation 4.7, values lower than 0.01 dB were found for both the SPL_{eq} method repeatability and method reproducibility contributions. For the headworn microphone Mipro MU-55HN, $u(SPL_{\mathrm{i,repr}})$ was equal to 0.1 dB, while a negligible uncertainty contribution of 0.01 dB was found for $u(SPL_{\mathrm{i,repe}})$. The method repeatability and reproducibility contributions of SPL_{eq} and SPL_{mean} were obtained over a total number of 120 frames and both were lower than 0.01 dB, according to equations 4.5 and 4.7.

In the case of SPL_{i}, the method reproducibility contribution is higher for Voice Care than for the headworn microphone due to the estimation of the calibration function, which is affected by the different positions of the contact microphone on the phantom material of the phonatory system simulator. This contribution becomes negligible in the case of SPL_{eq} and SPL_{mean} for both the devices. Particularly, considering a speech of 2 minutes and with 30% of phonation time percentage, it becomes comparable between the devices, even in the case of different frames length. Similar considerations are also valid for the method repeatability contribution in the case of differences between measures. Note that the method reproducibility and repeatability contributions for SPL_{eq} and SPL_{mean} would become even more negligible in the case of longer speech monitoring, according to equations 4.5 and 4.7. The method reproducibility contribution is not negligible for SPL_{mode}, but it is the smallest contribution for both the devices. The uncertainty contribution due to method repeatability for differences between measures does not result negligible only in the case of Voice Care, because it is based on the repeatability of the calibration function, as in the case of SPL_{i}.

Source reproducibility

The intra-speaker variability for SPL_{eq}, SPL_{mean} and SPL_{mode}, $u(SPL_{par,reps})$, estimated from the signals acquired with the Voice Care device was 0.8 dB, 0.6 dB and 1.5 dB, respectively, as reported in Chapter 3, paragraph 3.3.1. For the same SPL_{par} estimated with the headworn microphone $u(SPL_{par,reps})$ was equal to 0.5 dB, 0.6 dB and 1.1 dB, respectively.

Expanded uncertainty of speech sound pressure level and speech sound pressure level difference

An expanded uncertainty $U(SPL_i)$ equal to 2.6 dB and 1.4 dB was found for Voice Care and Mipro MU-55HN, respectively. $U(SPL_{par})$ resulted in 2.9 dB, 2.7 dB and 4.0 dB for Voice Care, and in 1.7 dB, 1.8 dB and 2.6 dB, for Mipro MU-55HN, in the case of absolute measure of SPL_{eq}, SPL_{mean} and SPL_{mode}, respectively. $U(SPL_{par})$ was 1.6 dB, 1.2 dB and 3.1 dB for Voice Care and 1.0 dB, 1.2 dB and 2.1 dB for Mipro MU-55HN, in the case of differences between the three SPL_{par}, respectively.

The expanded uncertainty for the instantaneous speech level, SPL_i, and the speech level parameters, SPL_{par}, is higher for Voice Care than for the headworn microphone. The instrumental uncertainty is the most influential contribution in the case of absolute measures, while the source reproducibility results to be the most significant contribution in the case of differences between measures. Generally, SPL_{eq} and SPL_{mean} expanded uncertainties are comparable for each device, both in the case of absolute values and differences between measures, with differences within 0.5 dB, while higher values have been obtained for SPL_{mode}. These findings agree with expectations, since for SPL_{eq} and SPL_{mean} the random contributions due to the method reproducibility and repeatability and to the source reproducibility entail the averaging among SPL_i values, while SPL_{mode} corresponds to a single SPL_i value, whose uncertainty is certainly higher.

As shown in Figure 9 of Švec *et al.* [18], the expanded uncertainty for SPL_{eq} and SPL_{mean} obtained from a contact-sensor-based device, were about 3 dB and 2 dB, respectively, over the same SPL range considered in this study. These values are rather comparable to those obtained in the present study with Voice Care, i.e. 2.9 dB and 2.7 dB, respectively. However, they did not consider the reproducibility contribution and their results are based on 27 subjects who read two different passages,

thus involving the inter-speaker variability. Instead, in the present study the results relate to individual SPL_{par}, i.e. they consider the intra-speaker variability, which implies higher source reproducibility contribution. When 27 subjects are involved, the source reproducibility contribution can be reduced to 0.5 dB, thus obtaining an expanded uncertainty of 2.6 dB for both SPL_{eq} and SPL_{mean}, according to the results described in Chapter 3.

It is important to consider the practical application of the results presented in this study that involve beyond researchers also practitioners in the field of vocal health, as speech therapists, ENT doctors and phoniatricians, and in the field of applied acoustics as acousticians or audio engineers. Whenever they have to compare two absolute speech SPL_{par} that imply the repositioning of Voice Care or Mipro MU-55HN, under changed conditions (e.g, different period of time, acoustics, subject, illness, age, ecc.), the SPL_{par} difference should be higher than $U(SPL_{\text{par}})$ values showed in Table 4.2 for absolute measures, in order to state that the changed condition significantly affect the speech production. On the other hand, when two speech SPL_{par} from the same subject have to be compared, without removing Voice Care or Mipro MU-55HN, their difference is significant if it is higher than $U(SPL_{\text{par}})$ values showed in Table 4.1 for differences between measures.

4.2 Investigation on the effects of very low and excessive reverberation in speech levels

Numerous studies have dealt so far with changing in speech production for talkers due to different acoustic environments, mostly focusing on the effect of noise [151, 111, 128, 152] or distance from the listeners [102, 153, 122, 101, 154] rather than on the effect of reverberation [111, 102, 58]. Reverberation was proved to influence voice production supporting talkers [111, 102] as well as increasing speech level towards an audience [155]. In spite of positive effects, excessive reverberation influences talkers making them use an erroneous vocal behaviour, which may be a cause of discomfort [156, 48] and a risk for vocal health [126], especially in the case of prolonged speech. Speaking is a very complex matter that involves many issues in addition to the mere presence of acoustical barriers, such as noise or reverberation. Speech modifications have been investigated at global, phonological and phonetic levels [131, 157] and can be determined by various combinations of talker characteristics, addressed listener

and listening environment [127]. Other factors of influence are the type of speech [158, 159] (e.g. read-, spontaneous-, simulated- and task-oriented- speech), and the speaking style [131, 157, 158] (e.g. clear speech or conversational speech). 'Clear speech' is a speaking style intrinsically more intelligible than ordinary, normally articulated conversational speech, that can be involuntary or deliberately produced in adaptation to a perturbed situation of communication or to a listener with reduced comprehension abilities. The way a talker addresses a listener can change according to the voice status [160, 146, 62], hearing sensitivity [161, 162], age [122] and gender [152, 122], mood and physical conditions [163, 164], speaking experience or training [165, 166]. Regardless of the listening environment, speech production is listener-oriented, since different interlocutor-related speeches (infant, foreigners, hearing-impaired persons, pets, machines) exist [127]. Type of listeners, presence of communicative intent [101, 155, 156, 158], eye contact [156] and familiarity [159] are therefore the main factors of influence. When background noise is present in the environment, a global increase of speech intensity occurs, leading to Lombard speech [128]. Such adaptation to the environment noisiness is highly variable from speaker to speaker, leading with a significant inter-speaker variability [152]. However, even in absence of masking noise, speech level increases can be observed at changing talker-to-listener distances [102, 153, 122, 101], perhaps as a form of compensation for perceived listener difficulties, and in communicative task [158]. Despite the extensive literature on the effects of noise on speech, to the Author' knowledge only few data have been published reporting details of the acoustic changes at global level that take place when a speaker modifies his vocal output while speaking in the presence of reverberation.

Brunskog *et al.* [55] and Pelegrín-García [58], investigated the effects of room acoustic parameters on the increase in the voice sound power level produced by six male speakers who held a lecture of about 5 minutes in six rooms with volume from 100 m^3 to 1900 m^3, and reverberation time in the range of 0.06 s (a 1000 m^3 anechoic chamber) to 1.53 s. Measurements of voice power level were based on speech signal acquired with a computer phone conversation headset, placed on the speaking subjects at about 3 cm from the mouth. They proposed a new objective parameter, namely the *room gain* [58], which represents the gain produced at the speaker's ears by the reflections in the room. From the model proposed by Brunskog *et al.* [55] , it appears that a talker tends to speak louder in rooms with a low room

gain (the anechoic chamber) and softer in rooms with a high room gain, which exhibits a greater support to voice production due to the room's reflections.

Pelegrín-García *et al.* [102] analyzed the effect of the acoustical environment on the natural speech evoked to describe a map [159]. They involved 13 male talkers aged between 23 and 40 years, addressing a listener at doubling communication distances (double distances increasing from 1.5 m to 12 m), in absence of background noise. They considered very different acoustic environments, among which an anechoic room and a reverberation room with a reverberation time averaged between 500 Hz and 1 kHz ($T30_{0.5\div1\text{kHz,occ}}$) of 0.04 s and 5.38 s, respectively. They measured the room gain at the talker position, from the mouth-to-ears impulse responses, which resulted to be 0.01 dB in the anechoic room and 0.77 dB in the reverberation room. The acoustic speech signal of each subject was picked up with a headworn microphone placed on the talker's cheek at a distance of 6 cm from the lips' edge. The length of the recordings varied between 1 and 2 min, depending on the map, which was different at each condition and that was administrated in random order for each subject, and the talker. At 6 m from the speaker's mouth (a distance which is representative of a lecturing scenario) they found, through a logarithmic regression model due to the differences among subjects, an increase in the mean sound power level (SWL) by 2.4 dB in anechoic chamber compared to reverberation chamber. The variability of the intercept and slope coefficients of such model resulted in a standard deviation of 2.74 dB and 0.76 dB/dd (i.e. per double distance), respectively, for mean SWL.

Cipriano *et al.* [167] investigated the relationships between room acoustics, background noise level and vocal effort of a speaker, the latter being expressed as equivalent speech sound pressure level at 1m from the speaker's mouth, in simulated classrooms of various volumes. The speakers, equipped with a headworn microphone, were found to adjust their vocal effort linearly with the voice support, i.e. the difference between the reflected sound level and the airborne direct sound level of the speaker's voice, at their own ears. The slope of this relationship, which was defined as room effect, was statistically significant and equal to -0.24 dB/dB in the case of the highest noise levels of 62 dB, thus supporting the increase of speech sound pressure level when voice support decreases, i.e. in a dead room. This finding was confirmed by Bottalico *et al.* [168], who monitored twenty subjects while reading a text in presence of speech babble noise (A-weighted equivalent sound pressure level, $L_{\text{A,eq}}$ of 62 dB) in anechoic and reverberant rooms with $T30_{0.5\div1\text{kHz,occ}}$ of

0.04 s and 2.37 s, respectively. An increase of about 1 dB was found in speech sound pressure level detected by an headworn microphone in anechoic room compared to reverberant room for both normal and loud speaking styles. In the studies presented so far, voice levels were measured at air-microphones specifically placed in a room at a given distance from the speaker's mouth.

As reported in Chapter 1 (pararaph 1.1.2), in recent studies contact-microphone-based vocal analyzers have been used in order to estimate vocal parameters from the skin vibration. These devices have been produced with the intent to perform long-term voice monitoring, since they have a negligible sensitivity to background noise [104]. The use of such devices was mainly investigated in in-field voice monitoring campaigns on teachers' vocal effort [155, 48, 21] (see Chapter 1, paragraph 1.1.3 and Chapter 2 for details). In particular, in all the referred studies the teachers' vocal effort was found to increase when classroom sound reflections increased, due to the contemporary increase of background noise level. The opposite behaviour of teachers' voice level compared to the above mentioned in-laboratory studies is hence mainly attributable to the background noise increase rather than to the sound reverberation increase.

In summary, there have been many studies reporting changing in voice intensity in presence of noise and at different communication distance, but only few compared only the effect of very low and very high reverberation time, in absence of noise. The present work investigates the variations of speech intensity in a group of subjects while speaking in semi-anechoic and reverberant rooms, with spontaneous speaking styles that are common in everyday life. Two types of spontaneous speech, namely a free monologue and the description of a map, have been addressed with a communicative intent from some speakers to a listener at a fixed distance of 6 m. Measurements were carried out with the commonly used headworn air-microphone and with a contact-based microphone vocal analyzer, which only detects the vocal strain without the influence of other acoustic artefacts.

4.2.1 Method

Experiments were carried out in the semi-anechoic and reverberant rooms of the National Institute of Metrological Research (I.N.Ri.M.) in Turin (Italy). The room volumes consisted in 384 m^3 and 294 m^3, respectively. The reverberation time value,

averaged in the frequency range between 0.5÷2 kHz was measured under empty rooms condition and it was equal to 0.11 s (SD 0.01) and 7.38 s (SD 1.61) for the semi-anechoic room and reverberant room, respectively. The overall equivalent A-weighted background noise level measured over a period of 5 minutes was equal to 24.5 dB and 30.3 dB, respectively.

Subjects and experiment instructions

The subjects involved in the study were asked to perform the spontaneous speeches being equipped with different microphones and devices, either together or separately. A total number of 38 speakers, (25 males, 13 females) was equipped with the Voice Care device, and a total number of 57 speakers (27 males, 30 females) was equipped with the headworn microphone. All subjects were either master or PhD students of Politecnico di Torino. Only native Italian speakers aged between 20 and 30 years were recruited. All of them did not have any severe visual impairment or any relevant vocal disorders, based on self-reports. Before starting the experiments, each speaker was asked to perform an audiometric screening test according to the procedure suggested by the iPad-based application titled uHear [143, 144], which provides a hearing sensitivity evaluation per frequency band (from 0.5 kHz to 6 kHz) with a level-based rating. Furthermore, the subjects equipped with the Voice Care device were asked to perform a calibration procedure in the semi-anechoic room, as described in Chapter 1, paragraph 1.1.2. Table 4.3 shows the number of subjects who undertook the various experiments in both the semi-anechoic and reverberant rooms. In the majority of the cases each speaker ran the speech task wearing the two different devices at the same time.

Once the preliminary operations were completed, the speakers were asked to produce a continuous 5 minute-long free speech, with the aim of transmitting information on something they knew well (e.g. the research topic they dealt with, a recipe, the rules of a game, the path from their house to the workplace), while standing 6 m away from a young female listener, sat on-axis in front of them such as to enable eye-contact. The listener, aged 24 years and self-reported normal hearing, had to take note on what the speakers said. Both the speakers and the listener were placed more than 1 m away from the boundary surfaces of the empty rooms.

The choice of making the subjects speak freely about a topic they knew well was related to the fact that this was considered the best way of making them express in a

normal speech manner. Reading or acting would have implied an inflection or an unnatural rhythm, so the vocal parameters would probably have been influenced by subjective and style factors rather than only by room acoustics [159].

In order to evoke another form of natural speech with a very specific communication intent [101], part of the subjects were also asked to describe a map. The map contained 12 landmarks (e.g., "school bus," "shop,", "yacht club"), a starting and an ending point marks, and a dashed line representing the path connecting these two points. Following the same procedure reported in Anderson *et al.* [159], the speakers were instructed to describe the route from the starting to the ending points, indicating the landmarks along the path (e.g., "go to the west until you find the yacht club"), while trying to enable visual-contact with the talker. The speaker had the task of making the listener draw the path correctly on a blank map containing all the items except the path and the ending mark. Cardinal points and 2.5 cm background square grid were provided on the map to facilitate the speaker-to-listener communication. Two maps were provided, one for each room, each sized 29.7 cm x 42.0 cm. The maps were printed on fabric and laid over a sound absorbing panel hung on a music stand, in front of the speaker's eyes, at a distance of 1.5 m, slightly to the left so that the listener's view was not perturbed. Each map description lasted from 2 to 3 minutes, depending on the speaker.

After explaining the speech tasks to the subject, the listener came back to her positions and indicated the speaker non-verbally when to start speaking. The listener gave no feedback to the speaker about the voice level perceived at her position, either verbally or non-verbally.

Subjects were asked to simultaneously utter the vowel /a/ and tap the ECM of Voice Care with their hands in order to produce sharp peaks on the signals acquired by the two microphones. The peaks were used at a synchronization scope in post-processing, with the aim of selecting the correspondent time histories of the two signals.

During the experiments in the semi-anechoic room, a reflection from the floor could have been occurred compared to a full anechoic room. In order to suppress this reflection thick sound absorbing panels were placed on the floor of the room.

Table 4.3 Number of subjects who undertook the experiments with Voice Care and the Mipro MU-55HN headworn microphone, for the speech tasks of free speech and describing a map. Distinction between female (F) and male (M) is also reported.

	Voice Care			Headworn microphone		
	F	M	Overall	F	M	Overall
Free speech	8	15	23	16	13	29
Describing a map	5	10	15	14	14	28
Overall	13	25	**38**	30	27	**57**

Speech level parameters

Mean speech sound pressure level, SPL_{mean}, mode, SPL_{mode}, and the overall equivalent sound pressure levels, SPL_{eq}, were calculated for each speech task and device, as described in paragraph 4.1.1. the Voice Care device estimated such sound pressure levels using *voiced frames* only based on a 30 ms frame length and applying the calibration function of the semi-anechoic room to both the monitorings in the two rooms. The sound power level, SWL, was also estimated for the headworn microphone, being aware that it is the most suitable parameter for a microphone in air, as it is directly related to the vocal strain [58, 55]. However, the contribution due to the reverberant field in the sound pressure level results negligible for such limited microphone-to-mouth distance, as showed below.

The relation between SWL and $SPL_{\text{eq,r}}$ in the reverberant room (r) can be expressed as $SWL = SPL_{\text{eq,r}} - (G_{\text{refl}} + G_{\text{dist}})$. The correction factor, G_{refl}, is due to the increase of the SPL_{eq} at the headworn microphone due to reflections in the reverberant room compared to the semi-anechoic room (sa) with the absorbing floor setting. For this measurements the 4128 B & K HATS emitting ICRA noise, equipped with the headworn microphone, was used and G_{refl} was determined as G_{refl}=$SPL_{\text{eq,r,HATS}}$–$SPL_{\text{eq,sa,HATS}}$. An average value of G_{refl} based on eight measures resulted equal to 0.34 dB (SD 0.05). Such a low value is expected since, as underlined by Brunskog *et al.*[55], the microphone was close enough to the source so that the direct field was predominant with respect to the reverberant field.

The correction factor G_{dist} depends on the source-receiver distance and source directivity. It was determined by performing sound power level measurements in the reverberant room with a HATS simulator, SWL_{HATS}, in a similar way as described by Brunskog *et al.* [55]. In particular, the 4128 B & K HATS was

placed in the reverberant chamber equipped with the headworn microphone; an ICRA noise signal was fed to the loudspeaker and measured simultaneously by the headworn microphone and by calibrated 1/2" microphone, B & K type 4943, located in the reverberant field of the room, according to the sound power level standard measurements ISO 3741 [169]. The correction factor has then been obtained as G_{dist}=$SPL_{\text{eq,r,HATS}}$−$(G_{\text{refl}} + SWL_{\text{HATS}})$ and resulted equal 23.3 dB (SD 0.05) based on eight measurements.

By assuming that all the speakers had the same directivity, equal to that of the HATS, the SWL of each subject in the reverberant room was finally estimated as described before, subtracting to each SPL_{eq} the two correction factors G_{dist} and G_{refl}. The SWL of each subject in the semi-anechoic room was instead obtained by subtracting to each SPL_{eq} only the correction factor G_{dist}. Since the Voice Care acquisition method is not affected by the environmental acoustic conditions, source directivity and microphone distance, SPL_{eq} is also representative of SWL for this device.

Statistical analyses

Different statistical analyses have been carried out with a MATLAB script and the results compared with IBM SPSS statistics package (version 21.0, Armonk, NY).

With the purpose of comparing the SPL parameters of the group of subjects in the two rooms, the non-parametric one-tailed Wilcoxon signed-rank test [149] has initially been applied. Assuming the dependency of the monitorings of the same subject in the two rooms, a paired list of samples has been considered for each SPL parameter. The speech parameters SPL_{eq}, SPL_{mean}, SPL_{mode} and SWL, which have been obtained in the two different rooms for each subject, constituted the two paired list samples. The test assessed the acceptance of the alternative hypothesis $H : M_{\text{sa}} > M_{\text{r}}$, where M_{sa} and M_{r} are the medians of each speech parameter list in the semi-anechoic and reverberant room, respectively (p-values lower than a significance level of 0.05).

SPL_{eq}, SPL_{mean}, and SPL_{mode} in the two rooms have been also compared through the estimation of their overall mean values among subjects for each task and device. The overall mean value of SWL has been also compared for the headworn microphone. Lately, the differences of the mean values between the two rooms ($\Delta SPL_{\text{sa}-\text{r}}$)

have been calculated for each parameter and they have been considered significant if their values exceed the respective expanded uncertainty for differences between speech levels, as explained in Chapter 3 in terms of inter-speaker variability, i.e. referred to a group of N speaker.

A further statistical investigation has been carried out applying the non-parametric Mann-Whitney U test [149] in its unilateral version only for SPL Voice Care data, which is directly related to the speech energy emitted by the speakers in the two rooms, being the contact microphone able to detect the vocal fold activity only. The test has been applied individually for each subject, considering his/her SPL distributions independent in the two rooms, as long as the speech evoked by the subject was different. The test assessed whether the two distributions in the two rooms of the same subject come from populations that are one stochastically larger than the other. It verifies the acceptance of the alternative hypothesis $H1 : M_{\mathrm{sa}} > M_{\mathrm{r}}$, where M_{sa} and M_{r} are the medians of the distributions of each speech parameter in the semi-anechoic and reverberant rooms, respectively (p-values lower than a significance level of 0.05).

4.2.2 Results

The results concern the comparison of speech sound pressure levels statistics between the different rooms, with the same device and the different tasks. The sound power level difference between the two rooms was also obtained and represents a measure of increased vocal effort due to the acoustic environment.

Due to the different characteristics of the used devices in terms of sample frequency, sensor positioning and unvoiced frames treatment, a direct comparison between the quantities estimated was not performed.

Voice Care

Table 4.4 shows the p-values of the one tailed Wilcoxon signed-rank test, which indicate a significant increase of SPL_{eq}, SPL_{mean}, SPL_{mode} for the group of speakers in the semi-anechoic room compared to the reverberant room, only in the case of describing a map. The same behaviour results not significant in the case of free speech.

Table 4.4 also shows the overall average value and standard deviation of the average of SPL_{eq}, SPL_{mean}, SPL_{mode} estimated with Voice Care at 16 cm from the speaker's mouth, in the semi-anechoic and reverberant rooms and the level differences between the two rooms ($\Delta SPL_{\text{sa}-\text{r}}$). Higher overall mean values of 2.1 dB, 1.8 dB and 2.4 dB are found in the semi-anechoic room compared to the reverberant room for SPL_{eq}, SPL_{mean}, SPL_{mode}, respectively, in the case of map description. These differences are significant as their values are higher than the respective expanded uncertainty for differences between speech levels of 1.4 dB for ΔSPL_{eq}, ΔSPL_{m}, and equal to 1.7 dB for ΔSPL_{mode} in the case of 15 subjects, according to paragraph 4.1.2. Since the Voice Care acquisition is not affected by the environmental acoustic conditions, source directivity and microphone distance, the 2.1 dB difference ΔSPL_{eq} in the two rooms can also be considered as representative of ΔSWL. In the case of free speech, the difference of SPL parameters in the two rooms were always lower than the respective uncertainty.

The outcome of the Mann-Whitney U-test related to the same speaker in the two rooms supports the finding of higher voice intensity in the semi-anechoic room compared to the reverberant room only in the case of map description. Table 4.6 shows that in the case of describing a map, 13 subjects out of 15 increased their voice level in the semi-anechoic chamber compared to the reverberant room, while only 10 out of 23 have the same behaviour in the case of free speech. The vocal behaviour of the subjects detected by Voice Care in the two rooms is also represented in Figures 4.2 and 4.3, where histograms of the SPL occurrences related to free speech and map description are shown for each subject.

The Voice Care results show a significant increase of about 2 dB in the sound pressure level parameters in the semi-anechoic room compared to the reverberant room for the map description. Assuming that ΔSPL_{eq} also represents the difference in ΔSWL between the two rooms, these results confirm the finding by Pelegrín-García *et al.* [102], who found an increase in the mean SWL by 2.4 dB in anechoic chamber compared to reverberation chamber with the same task but using an headworn microphone.

Such larger speech level increase when the speech task was that of clearly describing a map could be due to an higher motivation of speakers to make themselves understood, since the intent was to correctly explain directions to a listener who drew the path on a blank chart. This behaviour confirms the tendency of increasing speech

Table 4.4 Average value (upper cells) and standard deviation of the average (lower cells) of equivalent, SPL_{eq}, mean, SPL_{mean}, and mode, SPL_{mode}, sound pressure level (dB) estimated with Voice Care at 16 cm from the speaker's mouth, in the semi-anechoic (sa) and reverberant (r) rooms, and level differences between the two rooms ($\Delta SPL_{\mathrm{sa-r}}$). Results are shown for free speech and map description tasks. The p-values of the one-tailed Wilcoxon signed ranks test of the paired lists of parameters related to the two rooms are at the bottom. Values lower than a significance level of 0.05, reported in bold and italic style, indicate the acceptance of the alternative hypothesis $H : M_{\mathrm{sa}} > M_{\mathrm{r}}$, where M_{sa} and M_{r} are the medians of each SPL parameter list in the semi-anechoic and reverberant rooms, respectively.

Speech task	Subj.	SPL_{eq} sa	SPL_{eq} r	SPL_{mean} sa	SPL_{mean} r	SPL_{mode} sa	SPL_{mode} r	ΔSPL_{eq}	ΔSPL_{m}	$\Delta SPL_{\mathrm{mode}}$
Free speech	23	79.8 1.5	78.5 1.9	77.4 1.4	76.2 1.9	79.3 1.5	77.9 2.1	1.3	1.3	1.4
	p-value	*0.354*		*0.329*		*0.389*				
Map	15	79.0 2.1	77.0 1.8	78.7 2	76.8 1.6	82.0 2.5	79.6 1.8	2.1	1.8	2.4
	p-value	***0.004***		***0.007***		***0.004***				

level in clear speech with respect to conversational speech as found in literature [131, 127, 158]. The speech task of describing a map can be correctly configured as a clear speech task, since it also exhibited longer voicing periods compared to a conversational free speech accordingly to Astolfi *et al.* [126].

For both the rooms and the speech tasks, SPL_{eq} is greater than SPL_{mode}, as also found by Švec *et al.* [18] in the case of monologues evoked with vocal efforts similar to those of the present study. The mode is greater than the mean for all the scenarios, remarking therefore a non-normal distribution of the sound pressure level occurrences. The differences between SPL_{mode} and SPL_{mean} are greater in the case of describing a map than in the case of free speech and in the semi-anechoic room than in the reverberant room, hence supporting increased fatigue in the case of describing a map and speaking in a dead room [126].

Table 4.5 One-tailed Mann-Whitney U-test p-values on each couple of SPL distributions estimated for each male (M) or female (F) subject with Voice Care in the semi-anechoic (sa) and reverberant (r) rooms. Results are reported for both free speech and map description tasks. SPLs were obtained applying the calibration function of the semi-anechoic room to both the monitorings in the two rooms. Values lower than 0.05 are reported in bold and italic style and indicate the acceptance of the alternative hypothesis $H1 : M_{\mathrm{sa}} > M_{\mathrm{r}}$, where M_{sa} and M_{r} are the medians of SPL distributions in the semi-anechoic and reverberant rooms, respectively.

Subject	p-values	
	Free speech	Map
F01	1.000	-
F02	**0.000**	-
F03	1.000	**0.000**
F04	1.000	**0.000**
F05	**0.000**	**0.000**
F06	1.000	-
F07	1.000	1.000
F08	**0.000**	**0.000**
M01	**0.000**	-
M02	1.000	-
M03	**0.000**	-
M04	1.000	-
M05	**0.000**	**0.000**
M06	**0.000**	-
M07	**0.006**	**0.000**
M08	1.000	**0.000**
M09	1.000	**0.000**
M10	1.000	**0.000**
M11	**0.000**	**0.000**
M12	**0.000**	**0.000**
M13	1.000	**0.000**
M14	1.000	**0.000**
M15	1.000	1.000

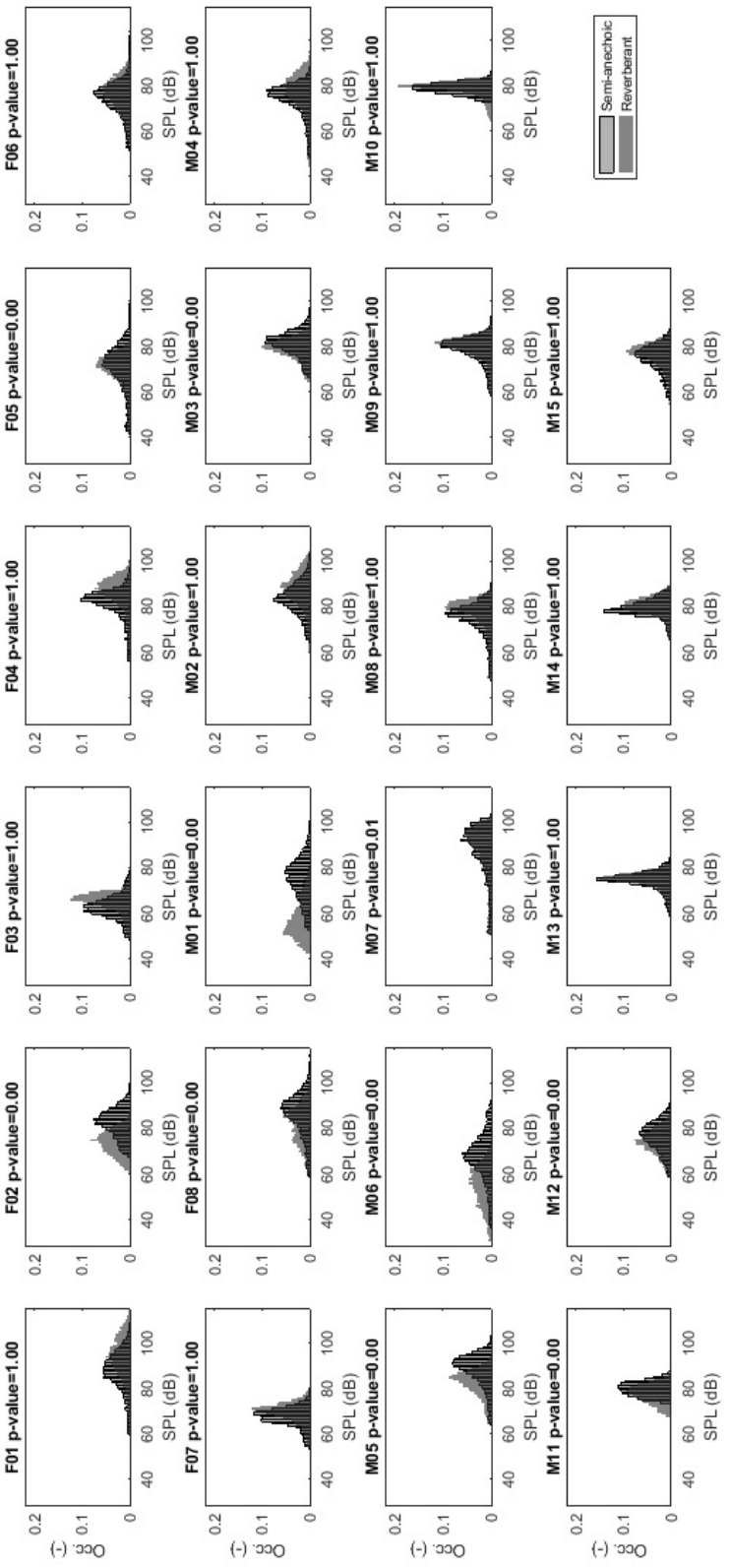

Fig. 4.2 Histograms of sound pressure level (SPL) occurrences related to 5 min of continuous free speech made by university students monitored using Voice Care in the semi-anechoic and reverberant rooms. *P*-values lower than 0.05 indicate that speakers rise their voice level in the semi-anechoic room compared to the reverberant room (10 out of 23 subjects).

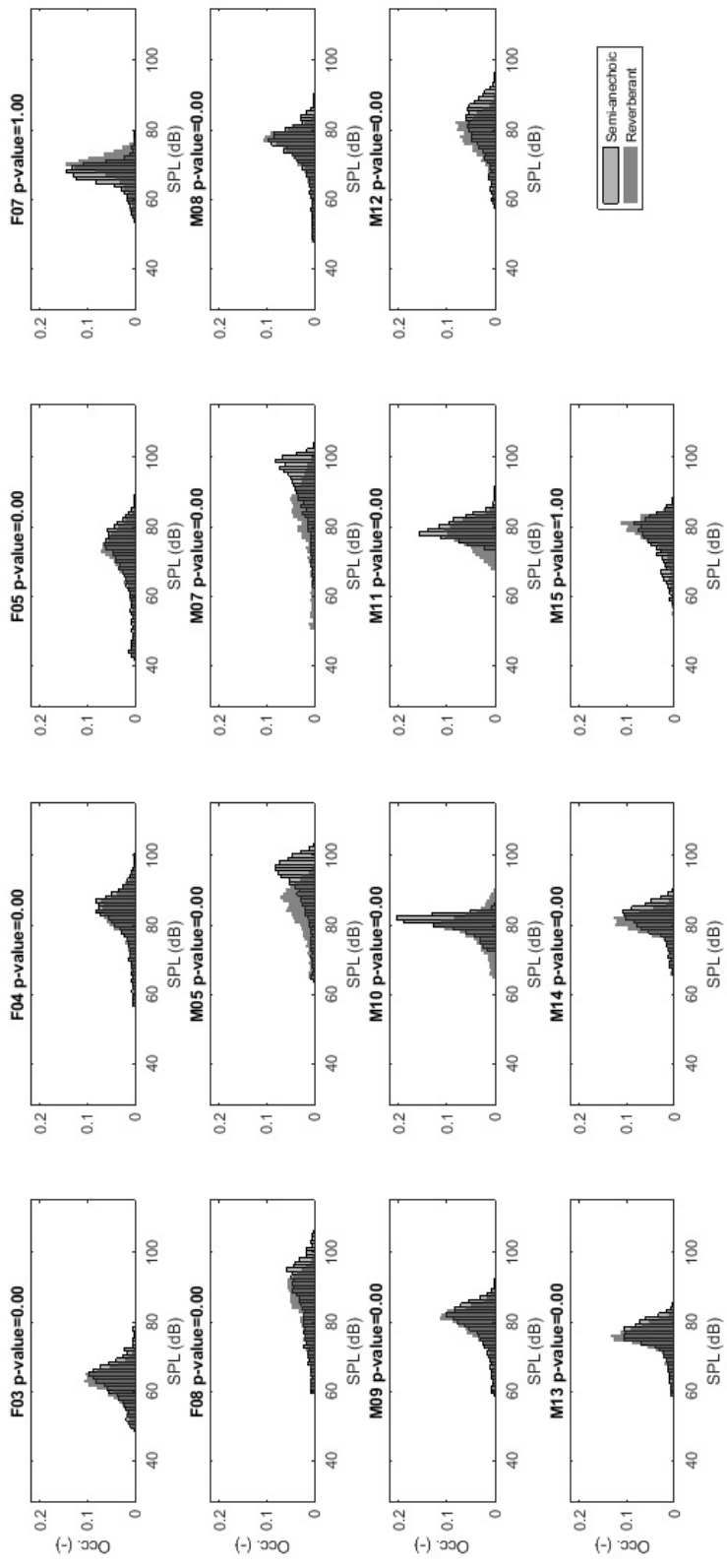

Fig. 4.3 Histograms of sound pressure level (SPL) occurrences related to speech samples in which a map was described by university students monitored using Voice Care in the semi-anechoic and reverberant rooms. *P*-values lower than 0.05 indicate that speakers rise their voice level in the semi-anechoic room compared to the reverberant room (13 out of 15 subjects).

Table 4.6 The same of Table 4.4 for data acquired with the headworn microphone Mipro MU55-HN.

Speech task	Subj.	SPL_{eq} sa	SPL_{eq} r	SPL_{mean} sa	SPL_{mean} r	SPL_{mode} sa	SPL_{mode} r	SWL sa	SWL r	ΔSPL_{eq}	ΔSPL_{mean}	ΔSPL_{mode}	ΔSWL
Free speech	29	94.3 0.8	93.5 0.9	92.0 0.8	91.7 0.9	94.1 0.9	93.9 0.8	71.0 0.8	69.8 0.9	0.8	0.3	0.2	1.2
	p-value	*0.051*		*0.276*		*0.233*		***0.009***					
Map	28	94.7 0.9	94.3 0.9	89.8 0.9	90.9 0.8	94.8 1.0	95.2 0.9	71.3 0.9	70.7 0.9	0.3	-1.1	-0.4	0.7
	p-value	*0.414*		*0.970*		*0.768*		*0.193*					

Headworn microphone

Table 4.6 shows the *p*-values of the one tailed Wilcoxon signed-rank test, which indicate a significant increase of SWL for the group of speakers in the semi-anechoic room compared to the reverberant room, only in the case of free speech. The same table also shows the overall average of SPL_{eq}, SPL_{mean}, SPL_{mode} and SWL, estimated with the headworn microphone Mipro MU-55HN at about 2.5 cm from the speaker's mouth, in the semi-anechoic and the reverberant rooms of I.N.Ri.M., and the level differences between the rooms.

All the differences are lower than the respective expanded uncertainty for differences between speech levels of 1.9 dB for ΔSPL_{eq}, 1.8 dB for ΔSPL_{m}, and equal to 2.0 dB for ΔSPL_{mode} in the case of 28 subjects, i.e. the smaller sample in Table 4.6, according to paragraph 4.1.2. Cautiously, in the case of ΔSWL the uncertainty could be considered the same of ΔSPL_{eq}, from which ΔSWL is obtained, where the contributions of G_{refl} and G_{dist} are not considered.

As far as the headworn microphone Mipro MU-55HN is concerned, the absence of significant differences between the paired lists of SPL_{eq}, SPL_{mean}, and SPL_{mode} related to the two conditions of semi-anechoic and reverberant rooms in the case of both free speech and map description, can be ascribed to different causes depending on either the speech parameter or the environment. These causes can be recognized in Figure 3 (upper chart), which shows the histograms of the SPL occurrences related to speech samples in which a map was described by a university student monitored using Mipro MU-55HN in the semi-anechoic and reverberant rooms. Firstly, the large logging interval of 1 s used for the analyses, which has a poor resolution compared to the voice frame lengths of 30-60 ms, could be the cause

of poor discrimination of SPL_{mode} between the two rooms. Secondly, the quite high values of the background noise levels recorded in the speech pauses both in the semi-anechoic room and in the reverberant room, contribute to the lowering of SPL_{eq} and SPL_{mean} in speech recordings. This is mainly valid for the SPL_{mean}, since the low number of the background noise levels occurrences negligibly affect SPL_{eq} estimation. However, background noise recorded in the speech pauses in the reverberant room is quite higher (around 75 dB) than in the semi-anechoic room (around 60 dB), since it is a reverberant noise, due to the long speech sound tail that fills the speech frame gaps. In the semi-anechoic room the background noise is due to the internal noise of the measurement chain that is higher than the background noise of the room, as highlighted in Chapter 3. This internal noise is masked by reverberation noise in the reverberant room. This behaviour determines a decrease of SPL_{mean} values in semi-anechoic room to a great extent, thus not highlighting an higher vocal strain expected in this room compared to the reverberant room.

When considering a shorter frame length of 30 ms, comparable to the inter-syllabic pause [32, 33], as in Voice Care, all the previous findings are emphasized, as shown in Figure 4.4(lower chart). As far as the SPL_{mode} is concerned, a bias can occur in the mode estimation in the case of both the semi-anechoic and reverberant room, where the lowest SPL peak-level positioned in correspondence of the background noise can overcome the highest SPL peak-level that identifies the speech levels. This is shown in the lower chart of Figure 3 in the case of speech taken in the semi-anechoic room. Anyway, if only speech level occurrences are considered, SPL_{mode} will be higher in semi-anechoic room than in reverberant room.

In light of the considerations expressed above, the adoption of 30 ms logging interval is not encouraged, as it would have brought to lower values of both SPL_{eq} and SPL_{mean} due to the presence of many background noise occurrences recorded in the speech pauses. In particular, a comparison between SPL parameters derived from the SPL distributions reported in Figure 4.4, processed with two different logging intervals of 30 ms and 1 s, brought to SPL_{eq} of 96.9 dB and 98.4 dB in semi-anechoic room, respectively, to 93.5 dB and 95.2 dB, for SPL_{eq} in reverberant room, respectively, to 78.8 dB and 95.3 dB, for SPL_{mean} in semi-anechoic room, respectively, and to 86.3 dB and 93.7 dB, for SPL_{mean} in reverberant room, respectively. With 30-ms logging interval, all the SPL parameters are lower, as expected, than the ones obtained with 1 s logging interval, and the most affected parameter is SPL_{mean}. In particular, the differences between 30 ms and 1 s logging interval are -1.5 dB and

-1.7 dB, for SPL_{eq} in semi-anechoic and reverberant room, respectively, and -16.5 dB and -7.4 dB for SPL_{mean} in semi-anechoic and reverberant room, respectively. Note that the differences in SPL_{eq} are the same for SWL, since they are equal up to an additive constant.

In conclusion, 1 s logging interval can be considered better than 30 ms for speech SPL parameters estimation with headworn microphones, but the problem of background noise recordings in the speech pauses still persists, making microphones in air less appropriate than contact-based microphone devices. In the case of the headworn microphone, the effective changing in voice intensity should be based on the differences between SWL where the gain due to reverberation, G_{refl}, in the reverberant room has been cut off although it was found to be negligible. According with the Wilcoxon signed-rank test, a significant increase in the SWL occurred in the semi-anechoic room only for free speech, but when the difference between the average SWL values is considered, it was lower than the related expanded uncertainty, allowing for a not significant result. ΔSWL is equal to 1.2 dB only, and such a low value could be due to the noise influence underlined above in the case of the headworn microphone. Moreover, the lack of unequivocal results in the case of the headworn microphone also concerns its distance from the subject's lips, which is not necessarily stable since its thin arch can cause slight changes of the position of the microphone during the experiment. On the contrary, the contact microphone of Voice Care is attached at the jugular notch of the subject, thus keeping a fixed position during the experiment.

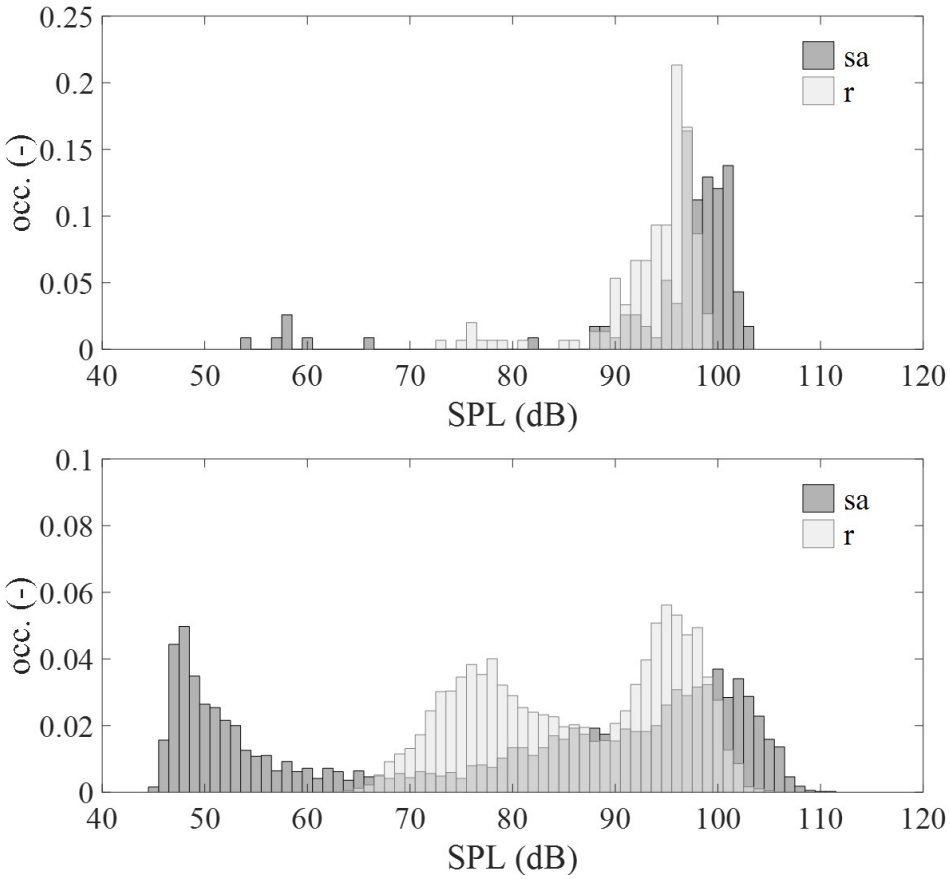

Fig. 4.4 Histograms of sound pressure level (SPL) occurrences related to speech samples in which a map was described by a university student monitored using Mipro MU-55HN in the semi-anechoic (dark grey) and reverberant rooms (light grey). Data refer to 1 s and 30 ms logging interval, in the upper chart and in the lower chart, respectively.

Chapter 5

Cepstral Peak Prominence Smoothed distribution in vowel as discriminator between healthy and dysphonic voice

This chapter partially reports material from:

1. A. Castellana, A. Carullo, S. Corbellini, A. Astolfi, M. Spadola Bisetti and J. Colombini, *Cepstral Peak Prominence Smoothed distribution as discriminator of vocal health in sustained vowel*, in Proc. IEEE I2MTC, Torino, Italy, May 22-25, 2017, pp. 552-557.

2. A. Castellana, A. Carullo, S. Corbellini, A. Astolfi, *Discriminating pathological voice from healthy voice using Cepstral Peak Prominence Smoothed distribution in sustained vowel*, IEEE Transactions on Instrumentation and Measurement, accepted.

The present chapter investigates Cepstral Peak Prominence Smoothed (CPPS) distributions in sustained vowel /a/ and their descriptive statistics as discriminators between healthy and unhealthy voices. Descriptive statistics different than the mean have been considered as possible candidate that could exhibit higher discrimination power. Signals acquired with two types of microphones have been included in the analysis, that are a headworn microphone and a contact Electret Condenser Microphone (ECM). The intra-speaker variability of CPPS parameters has been determined in repeated measures and the variability of the threshold values between healthy and unhealthy voices has been assessed by means of the Monte Carlo

method. Further investigations that are related to the identification of the main influence quantities of the cepstral parameters have been performed : among them, the fundamental frequency of the vocalization and the broadband noise superimposed to the signal have been taken into account. Eventually, the reliability of CPPS estimation with respect to the frequency content of the spectrum has been evaluated.

5.1 CPPS algorithm

A MATLAB (R2014b, version 8.4) script that is able to estimate the Cepstral Peak Prominence Smoothed according to Hillenbrand *et al.* [82] has been developed. Signals have been sampled at 22050 Hz and CPPS has been computed every 2 ms (frame) and using a 1024-point analysis window (46 ms). For each window, a series of operations lead to the cepstrum domain and then to the peak prominance estimation, after some smoothing processes. The following description summarizes the computational steps performed for each window of the vocal signal. Figure 5.1 shows the step-by-step outcome of these implementations in a analysis window. Starting from the signal in the time domain (fig. 5.1-a), the Fast Fourier Transform (FFT) algorithm has been employed on the Hamming-windowed signal in order to obtain the spectrum amplitude (fig. 5.1-b) and then the FFT algorithm has been used on the log power spectrum in order to reach the cepstrum domain (fig. 5.1-c). While the spectrum displays the energy of the frequency components within the signals through the harmonics, the cepstrum shows how regular the harmonic peaks in the spectrum are through the *rahmonics*. The cepstrum is in the time domain, as expected, that is here called *quefrency* and is usually expressed in milliseconds.

Two smoothing steps have been implemented on the obtained cepstra: first the smoothing in time that averages cepstra using a time-window of 14 ms (7 frames), i.e. each cepstrum has been replaced by the average of the current frame with the previous three frames and the following three frames; then the smoothing in cepstrum that averages the cepstral magnitude across quefrency with a 7-bin window, i.e. each cepstral magnitude is replaced by the average of the current bin with the previous and the following three bins. Figure 5.1-d represents the smoothed cepstrum obtained after the two-step procedure. Then, a linear regression line has been calculated in the quefrency to cepstral magnitude domain, between 1 ms and the maximum quefrency. The exclusion of the first millisecond from the regression line estimation is due to a

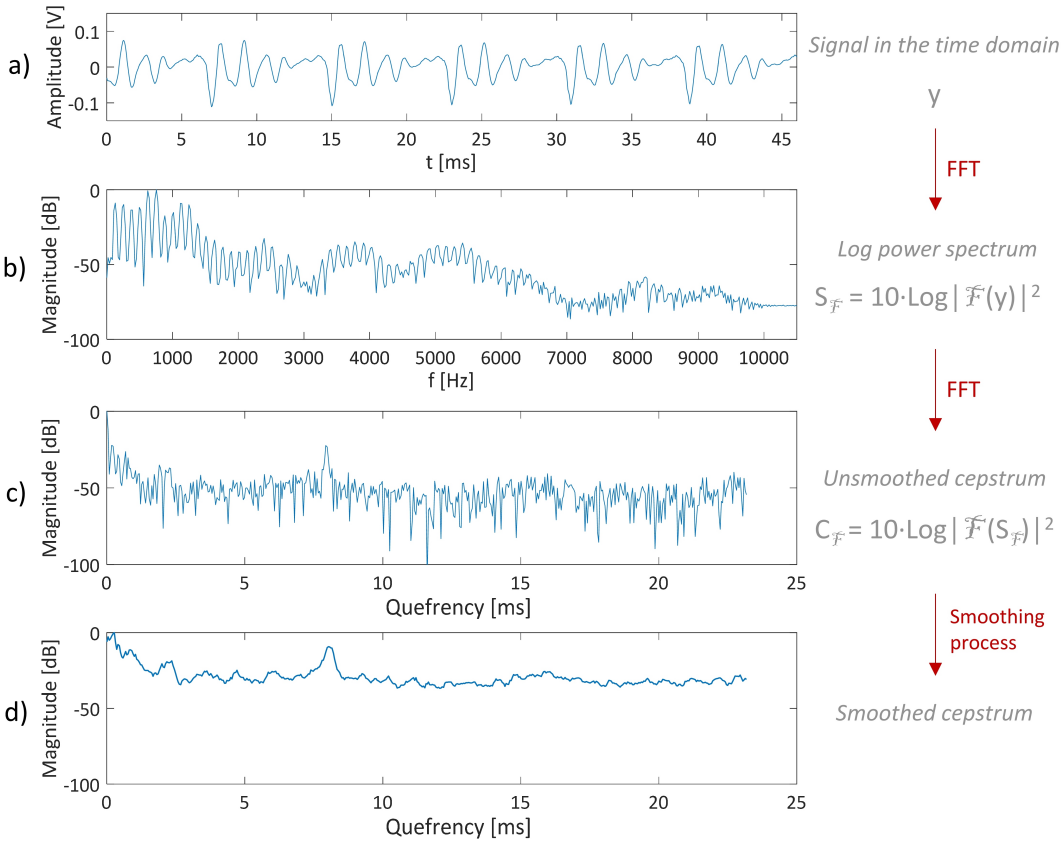

Fig. 5.1 From signal in time to smoothed cepstrum

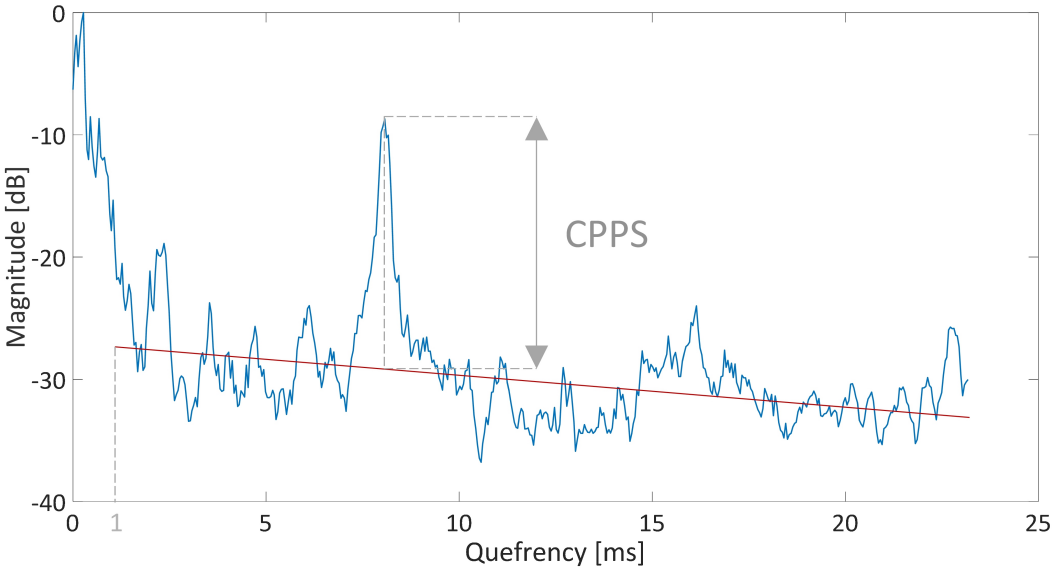

Fig. 5.2 CPPS calculation in smoothed cepstrum

property of the cepstrum of a voice signal that is highlighted in [170]: cepstrum at low quefrencies is more affected by the spectral envelope, which varies slowly, than by the spectrum periodicity.

The CPPS has been evaluated as the difference in dB between the peak in the cepstrum and the corresponding value at the same quefrency on the regression line, as shown in Figure 5.2. Since the quefrency at the cepstral peak generally corresponds to the inverse of the fundamental frequency, the cepstral peak have been looked for in the range between 3.3 ms (300 Hz) and 16.7 ms (60 Hz) in order to include the typical fundamental frequency range of female and male adults [171].

For each speech sample, the main outcomes of the algorithm are the CPPS occurrences that create individual CPPS distributions with a bin resolution of 0.1 dB. Figure 5.3 includes three examples of CPPS distribution obtained from a sustained vowel /a/ uttered by a person without voice problems and two people who suffered from voice disorders. For each CPPS distribution, different descriptive statistics have been estimated, which are able to characterize each distribution in location (mean, $CPPS_{\text{mean}}$, median, $CPPS_{\text{median}}$, 5th percentile, $CPPS_{5\text{prc}}$, and 95th percentile, $CPPS_{95\text{prc}}$), variance (standard deviation, $CPPS_{\text{std}}$, and the interval between the maximum and the minimum value, $CPPS_{\text{range}}$) and shape (kurtosis, $CPPS_{\text{kurt}}$, and skewness, $CPPS_{\text{skew}}$).

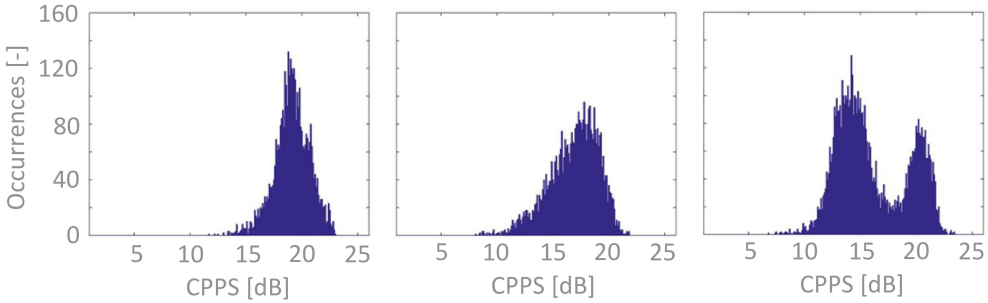

Fig. 5.3 Three examples of CPPS distributions obtained from the monitoring of a sustained vowel /a/ acquired with a microphone in air. From left to right: simmetric distribution with an higher mean for a healthy voice; a distribution with a negative skewness and a lower mean for a patological voice; a bimodal distribution for another pathological voice.

5.1.1 Comparison with existing software

Before addressing CPPS measures in healthy and pathological voice, the comparison between an existing software and the implemented MATLAB script has been performed. For this purpose, a subset of the collected database has been used. The voice recordings acquired with the headworn microphone as described in paragraph 5.2 from thirty patients with pathological voice (21 females and 9 males; age range: 21-81 years; mean: 58 years; standard deviation SD: 17.9 years) and thirty controls (20 females and 10 males; age range: 19-55 years; mean: 27.7 years; standard deviation SD: 10.2 years) have been included in this study. For each recording, CPPS values have beeh obtained using both the MATLAB algorithm shown in paragraph 5.1 and Hillenbrand software [172]. In particular, for each voice sample, $CPPS_{\text{mean}}$ from the MATLAB algorithm, $CPPS_{\text{mean}}$ *script*, has been compared with the unique output from the Hillenbrand software, $CPPS_{\text{mean}}$ *Hillenbrand*. On the command prompt of the software the same settings of the implemented algorithm were indicated before extracting CPPS values.

Figure 5.4 shows that CPPS measures in the two programs are strictly related, since the two lines representing the respective values in all the subjects follow the same path, with some exceptions. However, an offset in magnitude is present, in fact the blue line ($CPPS_{\text{mean}}$ *Hillenbrand*) is always under the red one ($CPPS_{\text{mean}}$ *script*) and it maintains a constant distance from the previous line: a mean value of 9.9 dB (standard deviation, SD: 3.2 dB) has been obtained for $CPPS_{\text{mean}}$ *Hillenbrand*, while

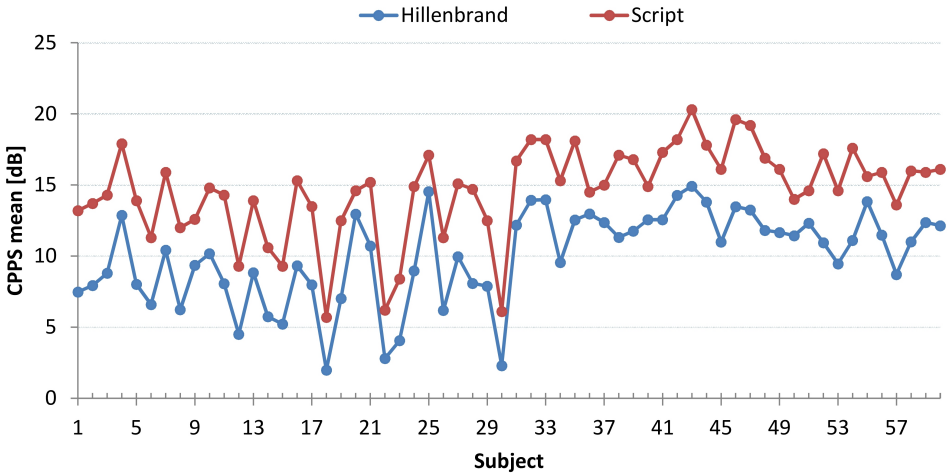

Fig. 5.4 Values of $CPPS_{mean}$ from the MATLAB script and CPPS from Hillenbrand software for each subject.

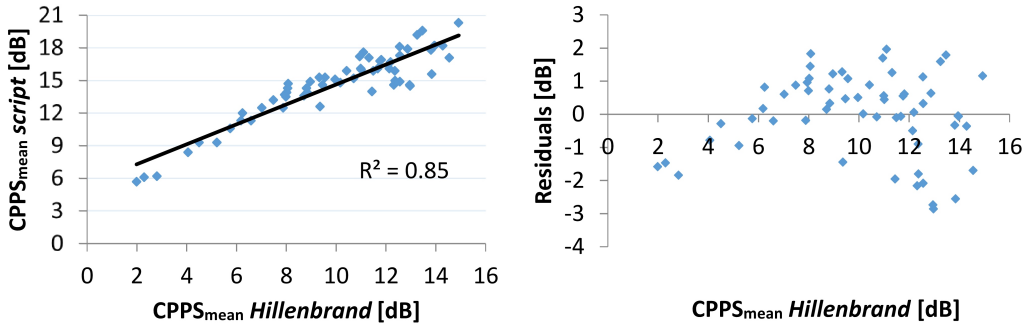

Fig. 5.5 On left side: scatter-plot with regression line between CPPS values from the MATLAB algorithm ($CPPS_{mean}$ *script*) and Hillenbrand software ($CPPS_{mean}$ *Hillenbrand*); on right side: plot of residuals in predicted $CPPS_{mean}$ *script* from actual observed $CPPS_{mean}$ *Hillenbrand*.

a mean value of of 14.6 dB (standard deviation, SD: 3.2 dB) has been obtained for $CPPS_{\text{mean}}$ *script*.

Figure 5.5 describes the strong relationship obtained between the two CPPS computations, which is characterized by a Pearson correlation coefficient of 0.93 (p-value < 0.001). The graph on the left shows the regression line with an Index of Determination (R^2) equal to 0.85, where R^2 indicates the amount of shared variation that is explained by the predictive regression model. The plot on the right reports the residuals, i.e. the errors not explained by the regression model, which are unbiased and homoscedastic. A standard error of estimates (SEE) equal to 1.2 dB resulted as an index of the average prediction error of $CPPS_{\text{mean}}$ *script* starting from $CPPS_{\text{mean}}$ *Hillenbrand*.

In summary, a high correlation was found between the two programs, thus validating the MATLAB script implemented as part of this research. Similar results have been obtained in a recent study [173], where CPPS from Analysis of Dysphonia in Speech and Voice (ADSV) and Praat were compared and a R^2 equal to 0.86 and 0.85 were found for Flemish and English vowels, respectively. Differences beetween two or more software products that are able to calculate CPPS may be due to methodological discrepancies in the spectral and cepstral procedures: in the case of Hillenbrand software and the MATLAB script, the unclear computational aspects are mainly about the windowing functions, the quefrency interval of the regression line calculation and the method of estimating the regression line. The Hillenbrand software follows the CPPS computational procedure given in [82], but the before-mentioned issues are not specifically described. The reader should take in mind that an offset in magnitude of 4.6 dB (SD: 1.24) is present between $CPPS_{\text{mean}}$ *script* and $CPPS_{\text{mean}}$ *Hillenbrand*.

5.2 Data collection

5.2.1 Subjects

Forty-one voluntary patients, 30 females and 11 males, participated in this study (age range: 20-77 years; mean: 51 years; standard deviation SD: 18.1 years). Thirty-five healthy adults with normal voices, 12 females and 23 males, were also included in the experiment (age range: 21-58 years; mean: 29 years; SD: 11.1 years). A clinical

protocol was followed for all the participants, who were all native Italian speakers. Table 5.1 shows the otolaryngologic diagnoses in the patient group.

5.2.2 Procedure

The protocol was designed in order to avoid each step affecting the following one. The relevant steps of the procedure can be summarized as follows:

1. each participant was asked to vocalize the sustained vowel /a/ on a comfortable pitch and loudness until he/she had need to breathe again, while he/she worn a headworn microphone and a contact microphone simultaneously;
2. participants repeated the previous task other two times, waiting few seconds of silence between the repetitions
3. two otolaryngologists performed the clinical practice that included a careful case history, auditory-perceptual measures (GRBAS scale) and the video-laringoscopy examination.

Figure 5.6 displays two moments of the protocol, which are the vocalization and the videolaringoscopy examination. In this study, the overall grade G of dysphonia of the auditory-perceptual measures has been reported only, since it has been the solely rating performed in consensus between the two otolaryngologists. For this reason, such data have been used with the purpose of commenting the results and not as further outcome of the work.

The vowel /a/ was selected as speech material due to its large use in acoustic analysis of voice and the duration of each phonation was always longer than 2 s, as recommended by Coleman [174].

5.2.3 Equipment for recording procedure

The voice recordings were performed in a quiet room of the Otolaringology department at the University Hospital "Città della Salute e della Scienza di Torino". The A-weighted equivalent background noise level in the room was measured with a calibrated class-1 sound level meter (NTi Audio XL2) over a period of 5 minutes

Table 5.1 Diagnoses for the patient group.

Organic dysphonia	**Patients**
Cyst	8
Edema	10
Sulcus vocalis	3
Polyp	4
Chronic laryngitis	4
Vocal fold hypostenia	3
Vocal fold paresis	2
Vocal fold nodul	2
Neurological disorder	3
Post-surgery dysphonia	2

Fig. 5.6 From left to right: a participant uttering the sustained vowel, while wearing both the microphone in air and the contact sensor; one otolaryngologist performing the video-laringoscopy examination.

in four different days, obtaining the average value of 50.0 dB (SD = 2.0 dB). The background noise level is 10 dB lower than the mean lowest A-weighted levels of 39 dB (60 dB) and 44 dB (65 dB) at 30 cm (2.5 cm), which were respectively found in healthy males and females producing their softest possible vowel [147]. Pathological voice tends to be softer than healthy voice, but in our experiment subjects were asked to read aloud, so an acceptable Signal-to-Noise Ratio was kept. Before performing

Table 5.2 Number of subjects who undertook the experiments with the Mipro MU-55HN headworn microphone and the ECM AE38 contact microphone. Number of patients and controls and females (F) and males (M) are also reported.

	Mipro MU-55HN			ECM AE38		
	F	M	Overall	F	M	Overall
Patients	30	11	41	28	6	34
Controls	12	23	35	12	23	35
Overall	42	34	76	40	29	69

the tasks described in steps (1) and (2), subjects worn the two microphones, which were:

- an omni-directional headworn microphone Mipro MU-55HN, which was placed at a distance of about 2.5 cm from the lips' edges of the talker, slightly to the side of the mouth. The microphone, which exhibits a flatness of ±3 dB in the range from 40 Hz to 20 kHz, was connected to a body-pack transmitter ACT-30T, which transmits to a wireless system Mipro ACT 311. The output signal of this system was recorded with an handy recorder ZOOM H1 (Zoom Corp., Tokyo, Japan), that use a sample rate of 44.1 kSa/s and 16 bit of resolution;
- an Electret Condenser Microphone (ECM AE38 [Alan Electronics GmbH (Dreieich, Germany)]), which was fixed at the jugular notch of each talker by means of a surgical band. The microphone senses the skin vibrations induced by the vocal-fold activity and it was connected to the handy recorder ROLAND R05 (Roland Corp., Milano, Italy), that samples the signal at a rate of 44.1 kSa/s using 16 bit of resolution.

Table 5.2 shows the details related to the subjects who performed the experimental task with the two microphones.

5.3 Analyses

Data were transferred from the handy recorders to a Personal Computer in order to be post-processed. First, the phonation interval from 1 s to 6 s has been selected

for each sustained vowel, using the software Adobe Audition (version 3.0). Then, CPPS has been estimated following the procedure described 5.1. Figure 5.7 shows the main computational steps in a single analysis window for a healthy voice and a severely pathological one: evident differences are present in the time domain of the signal, where the periodicity of the vowel is difficult to find in the sample of the patient; moreover, the spectrum of the healthy sample has separated harmonic peaks up to 4 kHz, while such a regularity is lost in the spectrum of the other voice sample. CPPS value represents such difference in regularity of the two spectra, which has a lower value for the pathological voice.

5.3.1 CPPS parameters in healthy and unhealthy voices

The two-tailed Mann-Whitney U-test [149] has been used to investigate statistical differences between each coupled list of descriptive statistics related to the patient group and the control subjects. It is a non-parametric test that refers to independent samples: the null hypothesis (H0) states that MD = 0, where MD is the median of the population of the differences between the sample data for patients and controls. When the null hypothesis is accepted, the two lists of values seem to come from the same population, i.e. it is not possible to distinguish healthy and unhealthy samples. The one-sample Kolmogorov-Smirnov test has been performed to verify that data in each list are not normally distributed, with the exception for the kurtosis values of CPPS distributions ($CPPS_{\text{kurt}}$) from patients. Such result allows the use of a non-parametric test for the analysis. The two above-mentioned tests have been performed using a MATLAB script.

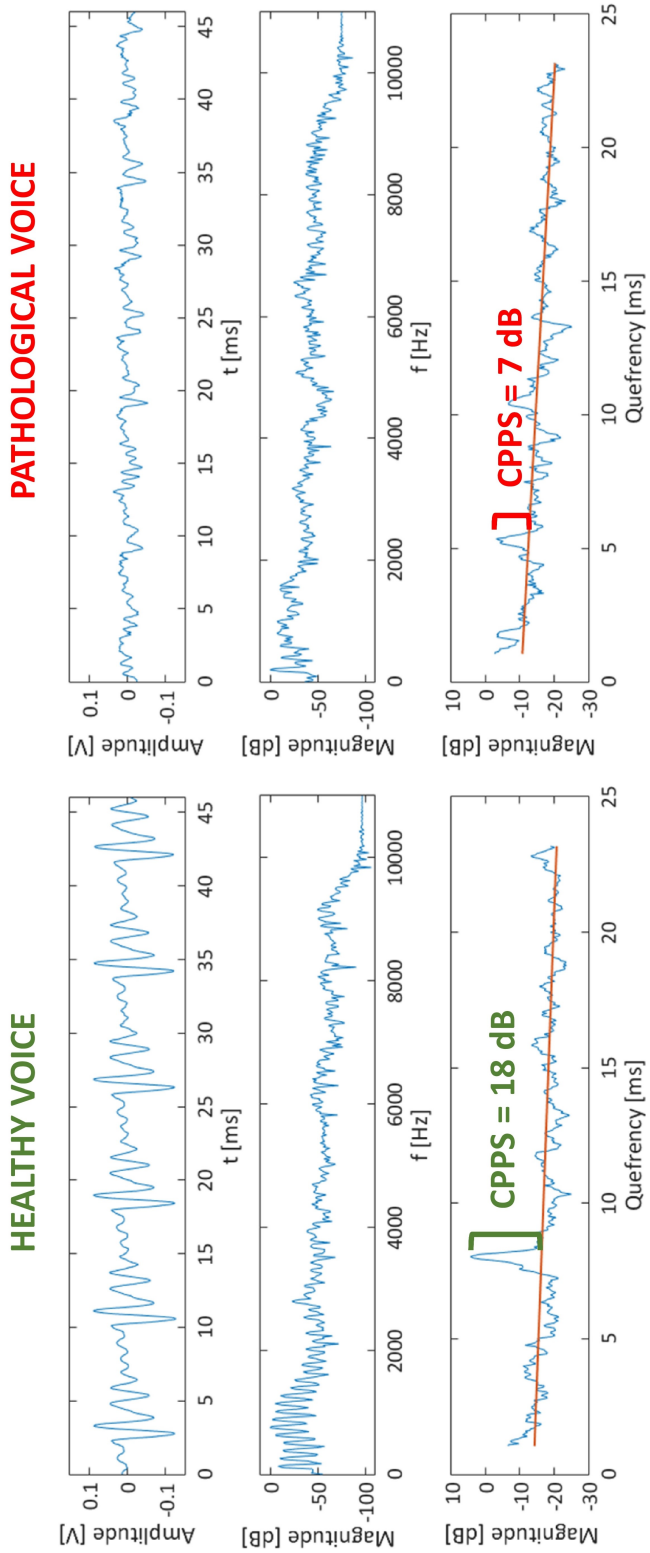

Fig. 5.7 From signal in time to CPPS of a sustained vowel /a/ of a healthy voice (on the left) and a pathological one (on the right).

5.3.2 Best logistic regression model

With the aim of investigating the effectiveness of the descriptive statistics for CPPS distribution as discriminators between dysphonic and healthy voices, a binary classification approach has been followed: a dichotomous variable, which has been coded as 0 or 1, has been given to each individual value of the descriptive statistics for CPPS distribution depending on the absence or the presence of dysphonia, respectively. The absence or the presence of the voice problem has been determined by the outcome of the videolaringoscopy examination. Then, a single-variable logistic regression model has been performed for each descriptive statistic and the best model was selected based on the highest Mc Fadden's R^2 and Area Under Curve (AUC) [175]. The Mc Fadden's R^2 characterizes the predictive power of a logistic regression model, while the area under the Receiver Operating Characteristic (ROC) curve describes the classification accuracy of the model. Area Under Curve (AUC) ranges from 0.5 to 1.0: an AUC near to 1 indicates a strong model's ability to separate those subjects with vocal disorders from those who have a healthy voice, while an AUC close to 0.5 means that the model has a poor capability to discriminate between the two groups.

Furthermore, the best threshold for the classification of healthy and pathological voices has been selected, observing a graph where sensitivity and specificity versus each possible threshold are plotted. Sensitivity is the true positive rate, i.e. the quota of people with voice problems who are correctly classified as positive. Specificity is the true negative rate, that is the percentage of subjects with healthy normal voice who are correctly identified as negative. The authors privileged a greater true positive rate (sensitivity) in selecting the best threshold, instead of taking the usual threshold that corresponds to the crossing point of sensitivity and specificity curves. All the analyses related to the logistic regression model has been performed using the statistical program RStudio (Version 0.99.489).

5.3.3 Intra-speaker variability

The repeatability of the descriptive statistics for CPPS distribution that have been included in the empirical fitted models has been investigated. Sixty-one subjects performed correctly the second task described in paragraph 5.2.2, while wearing both the headworn microphone and the ECM. For these participants, CPPS distributions have been calculated in the three repetitions of the sustained vowel /a/.

5.3.4 Monte Carlo method

The uncertainty estimation of the threshold values obtained for each logistic model has been assessed using the Monte Carlo method. First, the best fitting distribution for the lists of CPPS parameters that were included in the models has been determined through the Maximum Likelihood Estimation algorithm in MATLAB. This analysis has been performed for both healthy and pathological voices, including CPPS parameters from the three repetitions of the vowel for each subject. Then, 1000 trials of the Monte Carlo method have been repeated by randomly sampling 50 values from each fitted distribution. For each trial the best threshold of the logistic model has been determined, setting the equality between the sensitivity and the specificity obtained from the ROC analysis.

5.3.5 Influence quantities

The effects of fundamental frequency and broadband noise as influence quantities of the CPPS have been investigated by feeding the script that estimates the CPPS statistics with synthesized signals with well known characteristics. A set of vowels /a/ with the fundamental frequency in the range of 80 Hz to 260 Hz (frequency step of 20 Hz) has been synthetically generated using the software Sopran [176] with a sampling rate of 22050 $\mathrm{Sa/s}$. The selected frequencies cover both the typical female and male fundamental frequency range in sustained vowels of adults [171]. For each fundamental frequency, a 2 s long vowel has been created setting the first eight formants as pass-band filters with a Q factor of 20 and center frequencies of 580 Hz, 1.7 kHz, 2.9 kHz, 4.3 kHz, 5.4 kHz, 6.5 kHz, 7.7 kHz, 9.0 kHz. The Signal-to-Noise Ratio (SNR) of this set of vowels is of about 100 dB, which is mainly related to the quantization noise. Other two sets of vowels with the same frequency characteristics have been created adding two levels of random noise using MATLAB noise generator. A mean zero white Gaussian noise has been superimposed to the vowel signals setting the standard deviation in order to obtain SNR of 40 dB and 20 dB. For each fundamental frequency, CPPS distributions have been estimated by processing the 1 s long middle part of the vowel signal.

5.3.6 Frequency content of the spectrum

The 4 s middle part of a sustained vowel /a/ acquired with the headworn microphone from a control subject have been used in order to investigate the behaviour of CPPS distributions and their statistics with different frequency contents. Starting from the full spectrum bandwidth of the signal, that is of about 11 kHz, a 500 Hz frequency content has been cut away at a time and CPPS computation has been repeated for each step. This operation has been done down to a bandwidth of 1 kHz. Such analysis corresponds to an ideal low-pass filter with a cut-off frequency from 11 kHz down to 1 kHz and a step of 500 Hz.

5.4 Results and discussion

5.4.1 Microphone in air

The p-values obtained from the Two-tailed Mann-Whitney U-test of the lists of descriptive statistics related to the two groups of subjects were lower than 0.05, with the exception of skewness and kurtosis. These outcomes mean that the null hypothesis is rejected for most of CPPS parameters: CPPS distributions are significantly different in location, with an average value of 15.2 dB and 18.2 dB for $CPPS_{\text{mean}}$ in patients and controls, respectively, and in variance, with an average value of 1.9 dB and 1.3 dB for $CPPS_{\text{std}}$ in pathological and healthy voices, respectively.

Assuming the presence/absence of voice disorders as dependent variable, the best logistic regression model between healthy and unhealthy voice includes $CPPS_{5\text{prc}}$ as independent variable. The following formula defines the best empirical fitted model:

$$P(Unhealty) = \frac{e^{(28.8-1.93\cdot CPPS_{5\text{prc}})}}{1+e^{(28.8-1.93\cdot CPPS_{5\text{prc}})}} \tag{5.1}$$

where $P(Unhealthy)$ is the probability of having unhealthy voice, which ranges from zero to one. The negative coefficient of $CPPS_{5\text{prc}}$ shows that the probability to have unhealthy voice decreases as the $CPPS_{5\text{prc}}$ increases. A Mc Fadden's R^2 equal to 0.62 and an AUC of 0.95 of the model highlight that there is a clear separation between patients and controls: Fig. 5.8 shows the fitted values obtained for each subject and most of patients are in the upper part of the graph, where the probability of having

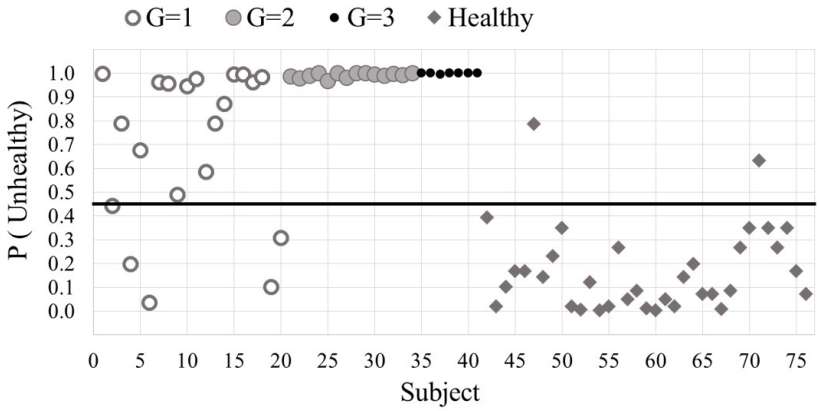

Fig. 5.8 Fitted values of the best logistic regression model, in terms of probability of having unhealthy voice, for vocalizations acquired with the headworn microphone Mipro MU-55HN. Circle points indicate the patient group (empty circles for the patients having a overall grade G of dysphonia equal to 1, gray circles for G=2 and black points for G=3); diamond points represent the control group. The bold line indicates the threshold value (0.44), which best separates patients and control subjects.

unhealthy voice is near to one, while most of controls have lower scores, near to zero. The best classification threshold was $P(Unhealthy) = 0.44$, that corresponds to 15.0 dB in terms of $CPPS_{5\text{prc}}$, with a sensitivity equal to 0.90 and a specificity of 0.94. As shown in Fig. 5.8, the four patients that are wrongly classified by the model have been judged with the lowest overall grade G of dysphonia.

The results on the repeatability of $CPPS_{5\text{prc}}$ are summarized in Fig. 5.9. For each subject, it shows the average values and the relative experimental standard deviations of the CPPS parameter in the three repetitions of the vowel /a/ acquired with the headworn microphone. Among the patient group, a clear separation between the first two grades G of dysphonia is not highlighted in the figure, while the three patients with G=3 show $CPPS_{5\text{prc}}$ lower than 8 dB. The average of the standard deviations of the $CPPS_{5\text{prc}}$ is equal to 0.8 dB for the patient group and 0.5 dB for the control group.

Fig. 5.9 also shows the threshold uncertainty, that is represented as a gray area around the $CPPS_{5\text{prc}}$ threshold. The probability density functions of the best-fitted distributions of $CPPS_{5\text{prc}}$ in pathological and healthy voices (bimodal and normal, respectively) have been used in a Monte Carlo simulation based on 1000 trials [177].

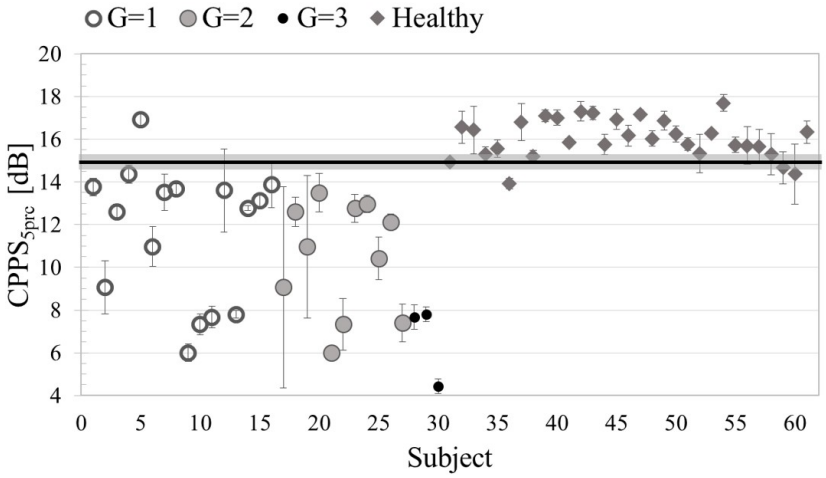

Fig. 5.9 Averaged values of $CPPS_{5\text{prc}}$ in the three repetitions of the vowel for each subject, acquired with the headworn microphone Mipro MU-55HN. Circle points indicate the patient group with different grades of dysphonia; diamond points represent the control group. Bars indicate the experimental standard deviation for each subject. The bold line indicates the threshold value (15.0 dB) and the gray area corresponds to its 95% confidence interval.

The output was a 95% confidence interval of the threshold equal to 0.7 dB, which constitutes the width of the gray area in Fig. 5.9.

5.4.2 Contact microphone

According to the outputs of the Two-tailed Mann-Whitney U-test, the lists of descriptive statistics for CPPS distributions related to the groups of patients and controls, who were recorded with the ECM, were significantly different in $CPPS_{\text{mean}}$, $CPPS_{\text{median}}$, $CPPS_{\text{std}}$, $CPPS_{\text{range}}$ and $CPPS_{5\text{prc}}$ (p-values < 0.05). As a consequence, CPPS distributions resulted significantly different in location, e.g. the average $CPPS_{\text{mean}}$ was equal to 18.0 dB for patients and 19.7 dB for controls, and in variance, e.g. the average $CPPS_{\text{std}}$ was equal to 1.7 dB and 0.9 dB for patients and controls, respectively.

The following formula describes the best empirical fitted logistic model for vowels acquired with ECM, which uses $CPPS_{\text{std}}$ as independent variable:

$$P(Unhealty) = \frac{e^{(-6.33+5.50\cdot CPPS_{\text{std}})}}{1+e^{(-6.33+5.50\cdot CPPS_{\text{std}})}} \tag{5.2}$$

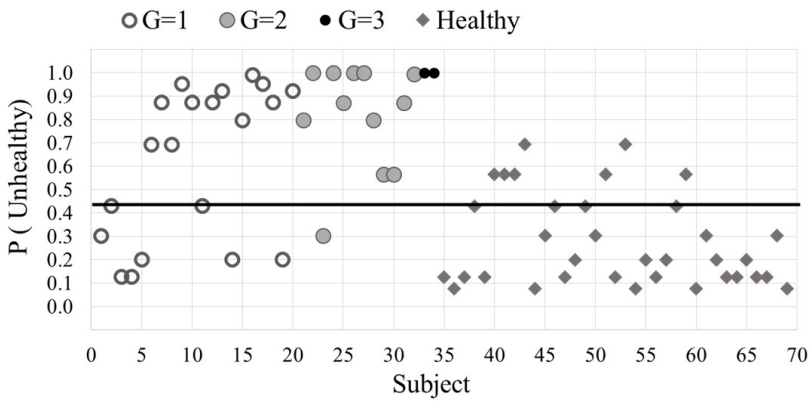

Fig. 5.10 The same of Fig. 5.8, for samples acquired with the contact microphone ECM AE38. The bold line indicates the selected threshold value, that is 0.43, which best separates patients and control subjects.

where $P(Unhealthy)$ is the probability of having unhealthy voice, which ranges from zero to one. The positive coefficient of $CPPS_{\text{std}}$ shows that the probability to have unhealthy voice increases as $CPPS_{\text{std}}$ increases. The empirical model has a moderate discrimination power with a Mc Fadden's R^2 equal to 0.38 and an AUC of 0.87: Fig. 5.10 shows that the fitted values of the two groups are not clearly separated. The best classification threshold is $P(Unhealthy) = 0.43$, that corresponds to 1.1 dB in terms of $CPPS_{\text{std}}$, with a sensitivity of 0.79 and a specificity of 0.69. Fig. 5.10 also shows that six out of seven patients that are wrongly classified by the model have been perceptually rated with the lowest overall grade G of dysphonia.

For each subject, the average values and the relative experimental standard deviations of $CPPS_{\text{std}}$ in the three repetitions of the vowel /a/ acquired with the ECM are reported in Fig. 5.11. One should note that patients rated with G=1 have lower $CPPS_{\text{std}}$ than those with G=2 and G=3. The average of the standard deviations of the $CPPS_{\text{std}}$ is equal to 0.3 dB for the patient group and 0.2 dB for the control group.

The same numerical procedure described in 5.4.1 has been implemented in order to estimate the threshold uncertainty, where a bimodal and a lognormal probability density functions have been used for pathological and healthy voices, respectively. The output was a 95% confidence interval of 0.2 dB. This interval is represented as a gray area around the $CPPS_{\text{std}}$ threshold in Fig.5.11.

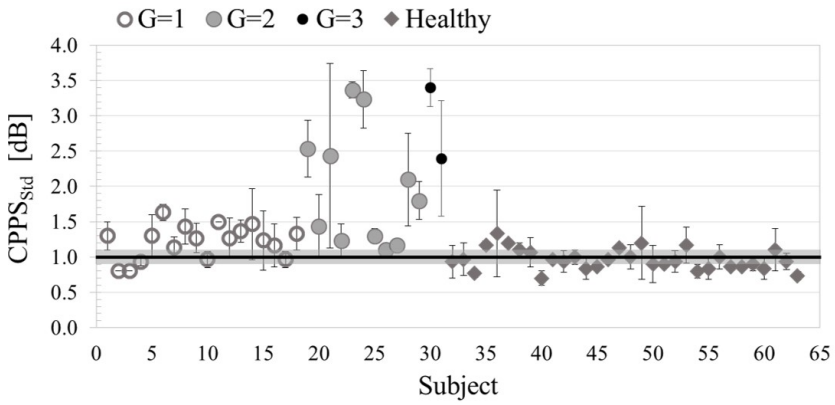

Fig. 5.11 Averaged values of $CPPS_{\text{std}}$ in the three repetitions of the vowel for each subject, acquired with the contact microphone ECM AE38. Circle points indicate the patient group with different grades of dysphonia; diamond points represent the control group. Bars indicate the experimental standard deviation for each subject. The bold line indicates the threshold value (1.1 dB) and the gray area corresponds to its 95% confidence interval.

5.4.3 Influence quantities: fundamental frequency and noise

Fig. 5.12 shows the behavior of $CPPS_{5\text{prc}}$ and $CPPS_{\text{std}}$ corresponding to the sets of vowels /a/ that have been synthesized according to the procedure described in the section 5.3.5.

The estimated $CPPS_{5\text{prc}}$ (red lines) shows a non monotonic behavior as the fundamental frequency increases for all of the three synthesized SNR levels. The standard deviation of the parameter $CPPS_{5\text{prc}}$ in the investigated frequency range resulted in 1.3 dB, 1.6 dB and 1.3 dB for SNR values equals to 100 dB, 40 dB and 20 dB, respectively. Hence the $CPPS_{5\text{prc}}$ shows a moderate dependence on the fundamental frequency, which is of the same order of magnitude of the estimated uncertainty of the discrimination threshold between healthy and unhealthy voices. However, the estimated standard deviation refers to a frequency range that includes both male and females voices, then lower variability is obtained by separating the two frequency ranges. In addition, it is possible to strongly reduce the observed variability by limiting the field of use of the fundamental frequency: from a practical point of view, this could be implemented by providing a reference frequency to the subject before he/she produces the sustained vowel. With respect to the SNR level, the three $CPPS_{5\text{prc}}$ curves are clearly separated: the one related to the highest SNR (100 dB) is above the other two curves, with an average value of 20.6 dB, while

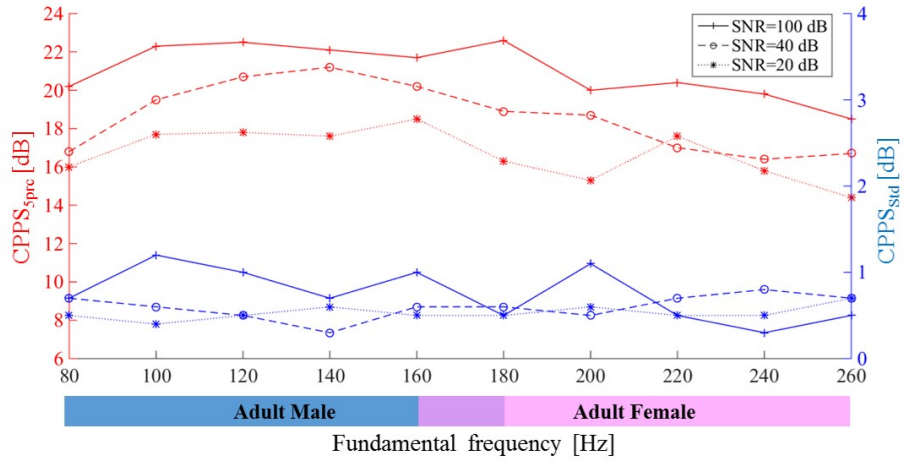

Fig. 5.12 Behavior of $CPPS_{5\text{prc}}$ (red lines) and $CPPS_{\text{std}}$ (blue lines) vs fundamental frequency, for three SNR levels (100 dB, 40 dB and 20 dB).

the one related to the noisiest signal (SNR of 20 dB) exhibits an average value of 16.3 dB. These findings confirm that the amplitude of the cepstral peak is dependent on the depth of the valleys between adjacent harmonics: higher the noise content in the spectrum shorter the height of the peak amplitude in the cepstrum [178, 179].

The parameter $CPPS_{\text{std}}$ (blue lines) vs the fundamental frequency is seemingly flat for the signals with SNR of 40 dB and 20 dB, while it exhibits an up-down trend when SNR is equal to 100 dB. Furthermore, $CPPS_{\text{std}}$ tends to rise as SNR increases: its average value in the investigated frequency range is 0.7 dB (standard deviation 0.3 dB) for $SNR = 100$ dB, 0.6 dB (s.d. 0.1 dB) for $SNR = 40$ dB and 0.5 dB (s.d. 0.1 dB) for $SNR = 20$ dB. This outcome proves that CPPS distributions have a higher variation when negligible noise is superimposed to the vocal signal.

One should note that the obtained values for the parameters $CPPS_{5\text{prc}}$ and $CPPS_{\text{std}}$ correspond to a healthy voice, since the former is higher than the identified threshold of 15.0 dB and the latter is lower than the threshold of 1.1 dB. This result, which is valid regardless of the effects of the investigated influence quantities, confirms the effectiveness of the proposed method, since synthesized vowels correspond to really healthy voices.

A further consideration can be made that is related to the differences of $CPPS_{5\text{prc}}$ and $CPPS_{\text{std}}$ between female and male typical fundamental frequency ranges. As shown in Fig. 5.12, adult male range is typically assumed from 80 Hz to 180 Hz,

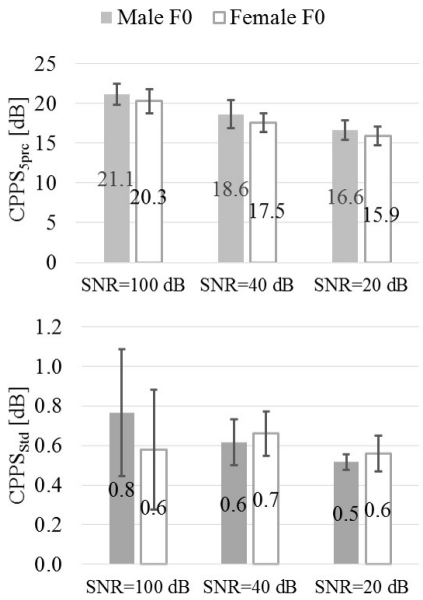

Fig. 5.13 Average values of $CPPS_{5\text{prc}}$ (upper part) and $CPPS_{\text{std}}$ (bottom part) in male and female frequency ranges; bars indicate the confidence interval obtained with a coverage factor $k = 2$.

while adult female fundamental frequency is in the range from 160 Hz to 260 Hz. As highlighted before, $CPPS_{5\text{prc}}$ curves have a slight down-trend as fundamental frequency increases. This seems confirmed by the results reported in the upper part of Fig. 5.13, since for the three investigated SNR levels the average of $CPPS_{5\text{prc}}$ is higher in the male range than in the female one. However, there is no significant difference between the two mean values of genders, since the standard deviations corresponding to the two frequency ranges overlap. The bottom part of Fig. 5.13 shows the behavior of $CPPS_{\text{std}}$ in male and female fundamental frequency ranges: also in this case, no significant differences have been found, even though the average $CPPS_{\text{std}}$ is higher in the male range than in the female one for $SNR = 100$ dB, while the opposite behaviour is observed for the other two SNR levels.

5.4.4 Frequency content of the spectrum

Fig. 5.14 shows how $CPPS_{5\text{prc}}$ (red line) and $CPPS_{\text{std}}$ (blue line) change when they are estimated from a healthy vowel /a/ whose spectrum has different frequency contents, starting from 11 kHz down to 1 kHz. Both the parameters have small

variations between 11 kHz and 5 kHz, then $CPPS_{5\text{prc}}$ increases reaching its maximum value for a frequency content of 3 kHz and it decreases again down to 1 kHz. The spectrum magnitude of the vowel under analysis, which is reported in the upper part of Fig. 5.14, highlights that the harmonic components between 5 kHz and 11 kHz have a limited energy content. In other words, these components contribute to the overall periodicity of the spectrum in a negligible way, so $CPPS_{5\text{prc}}$ keeps quite constant down to 5 kHz (the dotted black vertical line helps in reading the graphs). If instead the frequency content of the spectrum is limited to 3 kHz, sharp and clear harmonic components are deleted, which have an important role in the definition of the spectrum periodicity: for this reason $CPPS_{5\text{prc}}$ increases between 5 kHz and 3 kHz. Eventually, the parameter $CPPS_{5\text{prc}}$ decreases between 3 kHz and 1 kHz because of the limited number of harmonic components included in the spectrum. Differently from $CPPS_{5\text{prc}}$, $CPPS_{\text{std}}$ has a downward trend between 5 kHz and 3 kHz and it tends to have an up-down trend around a constant value again where the spectrum has a frequency content lower than 3 kHz. The reasons of such a change of behaviour can be found in the previous observations about the spectrum periodicity.

Fig. 5.14 also shows the frequency content of the signals acquired with the headworn microphone (MIC) and the ECM, which are respectively 10 kHz (vertical red dashed line) and 3.5 kHz (vertical blue dashed line). As we can observe in the graph at the bottom of the figure, the $CPPS_{5\text{prc}}$ has been estimated where its behaviour with the frequency content of the signal is almost stable, while $CPPS_{\text{std}}$, which is the best discrimination parameter for the ECM, has been estimated in the region where it shows a high dependence on the frequency content. This result suggests that the lower discrimination power that has been found for the ECM could be related to this effect. It is reasonable, since some voice qualities, e.g. breathiness, appears in the high frequency content of the spectrum. Then, such a voice acquired using a microphone with a limited bandwidth could be classified as healthy voice. However, further research that also involves pathological voices is necessary to assess this conclusion.

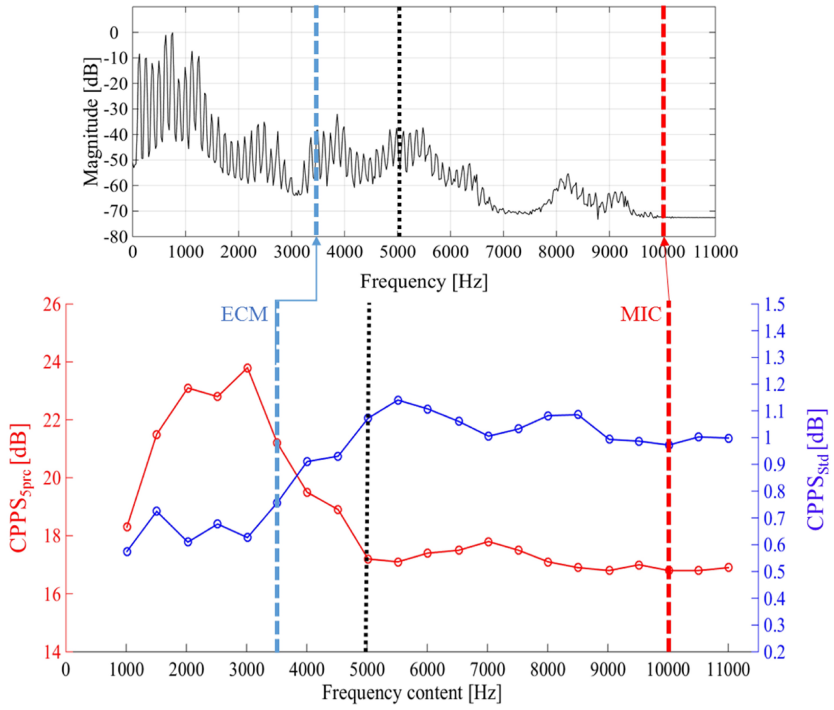

Fig. 5.14 (Bottom part) - Behaviour of $CPPS_{5\text{prc}}$ (red line) and $CPPS_{\text{std}}$ (blue line) vs frequency content of the spectrum. (Upper part) - Spectrum magnitude of the vowel under investigation, acquired with the headworn microphone. Vertical dashed lines correspond to the frequency content of signals acquired with the ECM (blue line) and with the headworn microphone (red line). Vertical dotted black lines helps in reading the graphs.

Chapter 6

Cepstral Peak Prominence Smoothed distribution in continuous speech

This chapter extends to the continuous speech the investigations on Cepstral Peak Prominence Smoothed (CPPS) distributions in sustained vowel /a/ described in Chapter 5. In the existing literature, despite earlier studies investigated time-based parameters, e.g. jitter and shimmer, in sentences [180, 181], it has been highlighted that they are only valid for sustained vowels produced with steady pitch and loudness, since any purposeful changes will be read as increases in vocal perturbation [80]. Spectral- and cepstral-based measures, instead, do not require cycle boundary detection, so they can be applied to continuous speech, which is more representative of everyday speaking patterns [81]. Several studies have highlighted the reliability of CPPS as measure of dysphonia in continuous speech [81, 182, 85], which best correlates with the perceptual evaluation of voice quality [84]. Moreover, CPPS represents the main contributor to the Acoustic Voice Quality Index (AVQI), which is a multivariate construct that yield a single number suitably correlated to overall dysphonia severity [183, 88, 184].

As highlighted in paragraph 1.2.1, all the existing studies on continuous speech only use microphones in air and consider as unique cepstral measure the mean of CPPS values and in some cases the standard deviation. This chapter includes three main studies on the topic. In **Study 1** other descriptive statistics for CPPS distribution in continuous speech as vocal health indexes have been investigated in both a microphone in air and a contact sensor, hypothesizing that other outcomes from

CPPS distribution may be useful in clinical practice, but also for patients feedback during everyday activities. Voice self-assessments and their relationship with CPPS parameters have been also detected. **Study 2** focuses on CPPS distributions in different voice qualities, when speeches are acquired using a microphone in air and two types of contact sensors, and on CPPS distributions as proof of outcomes after interventions as voice therapy and phonosurgery. Eventually, **Study 3** explores the variability of descriptive statistics for CPPS distribution within a healthy speaker and in a group of controls using two microphones in air and two contact sensors.

6.1 CPPS computation and comparison with existing software

CPPS computation in continuous speech includes a pre-processing step that account for removing unvoiced segments. A proper MATLAB algorithm has been implemented on a 1024-point analysis window, using the 60% of the signal root-mean-square obtained before removing unvoiced segments as threshold. Then, the same procedure described in Chapter 5, paragraph 5.1 has been followed.

The comparison between CPPS values from Hillenbrand software and the script used in this work has been carried out, using the all-voiced samples for feeding the programmes. In this way, the detection of pauses does not influence the final output.

Figure 6.1 shows that CPPS measures in the two programs are strictly related, since the red and blue lines follow the same trend point by point, with few exceptions. As for the sustained vowel, the two lines have a constant distance between them, that is an offset in magnitude of 4.5 dB (SD: 0.4 dB): a mean value of 7.3 dB (SD: 2.0 dB) has been obtained for the $CPPS_{\text{mean}}$ *Hillenbrand*, while a mean value of of 11.8 dB (SD: 2.2 dB) has been obtained for the $CPPS_{\text{mean}}$ *script*.

A strong relationship between the two CPPS computations was obtained, with a Pearson correlation coefficient of 0.99 (p-value<0.001). The graph on the left in Figure 6.1 shows the regression line between $CPPS_{\text{mean}}$ *Hillenbrand* and $CPPS_{\text{mean}}$ *script*, which is characterized by an Index of Determination (R^2) equal to 0.98, where R^2 indicates the amount of shared variation that is explained by the predictive regression model. The plot on the right represents the residuals, i.e. the errors not explained by the regression model, which are unbiased and homoscedastic. A

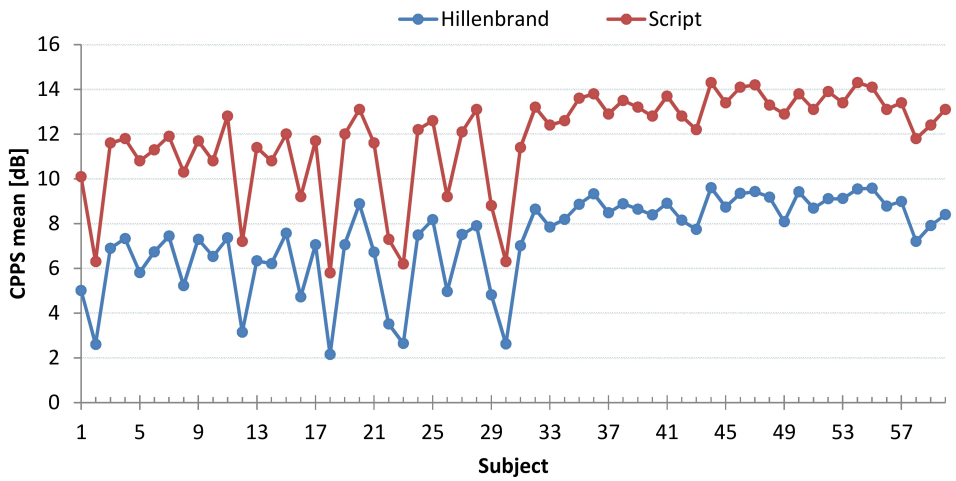

Fig. 6.1 Values of $CPPS_{\text{mean}}$ from the MATLAB script and CPPS from Hillenbrand software for each subject.

standard error of estimates (SEE) equal to 0.3 dB resulted as an index of the average prediction error of $CPPS_{\text{mean}}$ *script* starting from $CPPS_{\text{mean}}$ *Hillenbrand*. A recent study [173] has compared CPP in a sentence from *Analysis of Dysphonia in Speech and Voice*(ADSV) and *Praat* and comparable results have been obtained: Pearson correlation coefficients of 0.88 (R^2 of 0.77) and 0.96 (R^2 of 0.92) were found for Flemish and English speech, respectively, and a SEE equal to 0.6 dB and 0.7 dB were found as an index of the average prediction error of *ADSV* CPP starting from CPP *Praat* for the two languages, respectively. Furthermore, an offset in magnitude of about 13.5 dB was obtained between the two software estimations.

A similar investigation was carried out by Sauder *et al.* [185], who estimated CPPS in a sentence from the same software programs. Results were confirmed both with the same Pearson correlation coefficient of 0.88 and a comparable offset in magnitude between the two programmes of 13.8 dB. The outcomes obtained in the present work also confirm the results from Maryn and Weenink [186], who reported a very strong correlation between CPPS values obtained from *Praat* and *Speech Tool* in concatenated continuous speech samples and sustained vowels (Pearson correlation coefficients of 0.96 and R^2 of 0.92), despite clear differences in magnitude of CPPS estimated using the two programmes: e.g, a mean CPPS of 6.61 dB and 11.66 dB were obtained from *Speech Tool* and *Praat*, respectively.

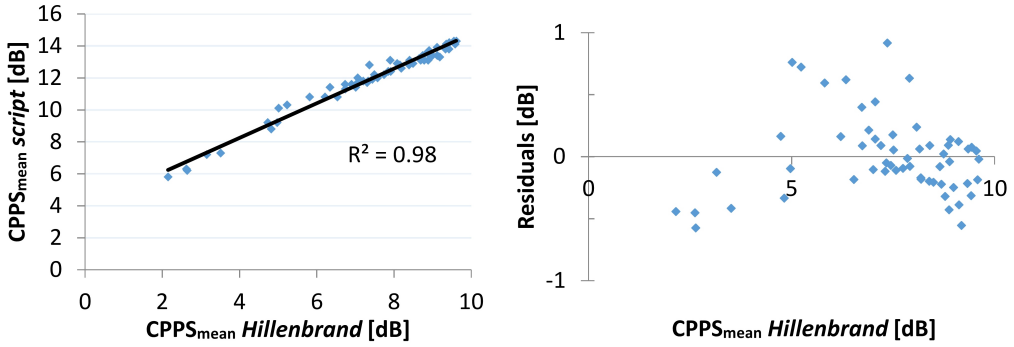

Fig. 6.2 On left: scatter-plot with regression line between CPPS values from the MATLAB algorithm ($CPPS_{\text{mean}}$ *script*) and Hillenbrand software ($CPPS_{\text{mean}}$ *Hillenbrand*); on right: plot of residuals in predicted $CPPS_{\text{mean}}$ *script* from actual observed $CPPS_{\text{mean}}$ *Hillenbrand*.

In summary, the CPPS algorithms implemented in the existing software and in the MATLAB script as a part of this research have methodological discrepancies in the spectral and cepstral procedures that lead to different CPPS outputs that are strongly correlated with each other. As already specified in Chapter 5, the windowing functions, the quefrency interval of the regression line calculation and the method of estimating the regression line are the mainly unclear computational aspects between Hillenbrand software and the MATLAB script. Since the Hillenbrand software follows the CPPS computational procedure given in [82], but the before-mentioned issues are not specifically described, it is difficult to compare CPPS values estimated from the software with the MATLAB script systematically. Moreover, as previously described from the cited references, comparisons between different software programs provide different outputs. For this reason, the reader should take in mind that an offset in magnitude of 4.5 dB (SD: 0.4 dB) is present between $CPPS_{\text{mean}}$ *script* and $CPPS_{\text{mean}}$ *Hillenbrand*.

6.2 Study 1: Cepstral Peak Prominence Smoothed distribution in continuous speech as vocal health indicator

6.2.1 Method

Seventy-two voluntary patients, 55 females and 17 males, participated in this study (age range: 20-82 years; age mean (M)=54; age standard deviation (SD)=17.2). Thirty-nine healthy adults with normal voices, 16 females and 23 males, were also included in the experiment (age range: 19-58 years; M=29.9; SD=11.7).

All participants were native Italian speakers. Primary disorders in the dysphonic group included edema (10), cyst (10), sulcus vocalis (6), polyp (5), chronic laryngitis (6), vocal fold hypostenia (7), post-surgery dysphonia (3), vocal fold paresis (8), vocal fold nodule (5), neurological disorder (6), functional dysphonia (6).

The protocol was designed in order to not allow each step influencing the following others and it can be summarized as follows:

1. each participant filled a self-assessment questionnaire, the Italian version of the Voice Activity And Participation Profile (Profilo di Attività e Partecipazione vocale, PAPV)[187];

2. each participant was asked to read aloud an Italian phonetically balanced passage, which is a short tale of 300 words. It took an average reading time of about 2 minutes. The reading text is the first reported in the appendix.

3. each participant was asked to make a continuous 2 minute-long free speech, with the aim of telling something they knew well, e.g. their last summer holidays or the path from their house to the workplace, while standing 2 m away from a young female listener, sat on-axis in front of them.

4. two otolaryngologists performed the clinical practice that included a careful case history, auditory-perceptual measures (GRBAS scale) and the video-laringoscopy examination. The two raters performed in consensus the overall grade G of dysphonia only, then only such a rating has been used for the discussion of the results in the present study.

The voice recordings were performed simultaneously using the omnidirectional headworn microphone Mipro MU-55HN and the Electret Condenser Microphone ECM AE38 (for further details see paragraph 5.2.3). For unavailability of one of the two devices, 66 out 72 patients performed the experiment using both the headworn microphone and the ECM, while 6 out 72 worn only the first. The wav files were down-sampled to 22.05 kSa/s and a CPPS distribution has been computed for each speech task, as described in paragraph 6.1. For each CPPS distribution, the following descriptive statistics were estimated, named as CPPS parameters: mean, $CPPS_{\text{mean}}$, median, $CPPS_{\text{median}}$, mode, $CPPS_{\text{mode}}$, 5th percentile, $CPPS_{5\text{prc}}$, 95th percentile, $CPPS_{95\text{prc}}$, standard deviation, $CPPS_{\text{std}}$, the interval between the maximum and the minimum value, $CPPS_{\text{range}}$, kurtosis, $CPPS_{\text{kurt}}$, and skewness, $CPPS_{\text{skew}}$.

6.2.2 Analyses and results

Different statistical analyses have been performed on the CPPS parameters obtained from the groups of patients and controls in order to investigate if they are significantly different in the two groups, and to explore their diagnostic precision and their agreement in the two speech materials. The same analyses have been repeated for both the devices.

CPPS parameters in healthy and unhealthy voices

Firstly, the two-tailed Mann-Whitney U-test, a non-parametric test based on independent samples, has been applied on each coupled list of descriptive statistics related to the group of patient and the control subjects (details in paragraph 5.3.1). Then, the efficacy of the CPPS parameters as discriminators between dysphonic and healthy voices has been investigated by means of a binary classification analysis, as described in paragraph 5.3.2. A single-variable logistic regression model has been performed for each descriptive statistic and the best model has been selected based on the highest Mc Fadden's R^2 and Area Under Curve (AUC). Moreover, the leave-one out classification accuracy has been also evaluated. The optimal threshold for the classification purpose has been selected accounting for the highest sensitivity and specificity in correspondence of each possible cut-off point, with a preference of having a greater sensitivity between the two. Sensitivity, that is the true positive rate, is the proportion of subjects with voice pathology who are correctly classified

as positive. Specificity, that is the true negative rate, is the percentage of people with healthy voice who are correctly identified as negative. The statistical program RStudio (Version 0.99.489) has been used for these analyses.

Table 6.1 shows the p-values of the Two-tailed Mann-Whitney U-test for the headworn microphone: they are lower than 0.05, which means H0 rejected, for all CPPS parameters with the exception of $CPPS_{\text{kurt}}$. These results, which are the same for both the speech materials, highlight that the coupled lists of all the CPPS parameters apart from $CPPS_{\text{kurt}}$ in healthy and unhealthy groups are significantly different. Such outcomes on the descriptive statistics for CPPS distributions reveal that also CPPS distributions themselves from healthy and dysphonic speakers are significantly different, both in central tendency and in variance. Table 6.1 also shows the performance of each model with a CPPS parameter at a time as independent variable and the presence/absence of voice disorders as dependent variable on classifying healthy and unhealthy voice. The best logistic regression model between healthy and pathological voice includes $CPPS_{95\text{prc}}$, which has the highest Mc Fadden's R^2 and AUC in both the speech materials. The respective values are also very similar: the Mc Fadden's R^2 is equal to 0.34 for the reading task and 0.33 for the free speech, while the AUC is 0.86 in both the cases. Such AUC value designates an excellent discrimination power between the two groups of speakers. The leave-one-out cross validation accuracy is comparable with other CPPS parameters, such as $CPPS_{\text{mean}}$ and $CPPS_{\text{median}}$, for both reading and free speech: it is equal to 77% and 73%, respectively.

The following formula defines the general best empirical fitted model for both reading and free speech tasks:

$$P(Unhealthy) = \frac{e^{(intercept - slope \cdot CPPS_{95\text{prc}})}}{1 + e^{(intercept - slope \cdot CPPS_{95\text{prc}})}} \tag{6.1}$$

where $P(Unhealthy)$ is the probability of having unhealthy voice, which ranges from zero to one.

Table 6.2 summarizes the best empirical logistic models related to reading and free speech acquired with the headworn microphone. They both include $CPPS_{95\text{prc}}$ as independent variable, which has a negative coefficient that indicates the increase of the probability of having unhealthy voice when $CPPS_{95\text{prc}}$ decreases. The best threshold selected from the reading model is 18.1 dB, while the best threshold for the

Table 6.1 Analysis results for each CPPS parameter related to the headworn microphone. Two-tailed Mann-Whitney U-test p-values: values lower than 0.05 are in bold and indicate the rejection of the null hypothesis. Logistic regression model: Mc Fadden's R^2 (Mc Fad.), Area Under Curve (AUC) and its relative 95% Confidence Interval (CI), leave-one out classification accuracy (acc.). The line in italic indicates the CPPS parameter included in the best logistic model.

CPPS param.	**Reading**				**Free speech**			
	U-test	Mc Fad.	AUC (CI)	Acc.	U-test	Mc Fad.	AUC (CI)	Acc.
$CPPS_{\text{mean}}$	**0.001**	0.30	0.84 (0.77-0.92)	77%	**0.001**	0.24	0.80 (0.72-0.88)	74%
$CPPS_{\text{median}}$	**0.001**	0.27	0.83 (0.76-0.91)	77%	**0.001**	0.23	0.79 (0.72-0.88)	74%
$CPPS_{\text{mode}}$	**0.001**	0.22	0.79 (0.71-0.88)	72%	**0.001**	0.16	0.79 (0.71-0.88)	71%
$CPPS_{\text{std}}$	**0.001**	0.14	0.74 (0.64-0.84)	73%	**0.001**	0.14	0.76 (0.67-0.86)	71%
$CPPS_{\text{range}}$	**0.001**	0.27	0.83 (0.76-0.91)	74%	**0.001**	0.19	0.80 (0.72-0.89)	72%
$CPPS_{5\text{prc}}$	**0.001**	0.17	0.78 (0.70-0.87)	74%	**0.001**	0.14	0.74 (0.65-0.83)	72%
$CPPS_{95\text{prc}}$	***0.001***	*0.34*	*0.86 (0.80-0.93)*	*77%*	***0.001***	*0.33*	*0.86 (0.79-0.93)*	*73%*
$CPPS_{\text{skew}}$	**0.001**	0.15	0.72 (0.63-0.82)	70%	**0.001**	0.13	0.70 (0.61-0.80)	69%
$CPPS_{\text{kurt}}$	0.685	0.01	0.52 (0.41-0.63)	65%	0.369	0.01	0.55 (0.44-0.67)	66%

Table 6.2 Best logistic models including $CPPS_{95\text{prc}}$, related to reading and free speech acquired with the headworn microphone Mipro MU-55HN. The threshold value and the respective sensitivity (sens.) and specificity (spec.) are also reported.

Speech material	**Model**	**Estimate**	**Threshold (dB)**	**Sens.**	**Spec.**
Reading	Intercept	27.5	18.1	0.82	0.77
	Slope	-1.5			
Free Speech	Intercept	25.3	17.9	0.78	0.74
	Slope	-1.4			

free speech model is 17.9 dB. The first threshold exhibits sensitivity and specificity of 0.82 and 0.77 respectively, which are higher than those of the latter one (0.78 and 0.74, respectively).

Figure 6.3 shows the fitted values obtained for each subject by the best logistic model on reading samples. Most of patients are in the upper part of the graph, where the probability of having unhealthy voice is near to one, while most of controls have lower scores, near to zero. Moreover, there is a partial agreement between the overall grade of dysphonia G and the probability of having unhealthy voice: all subjects with G=3 have fitted values close to 1, subjects with G=2 lie in the upper part of the graph (with an exception only), but subjects with G=1 are spread in the graph. Such an evidence means that patients who were perceptually labelled with the lower grade of dysphonia include both people who are able to compensate for

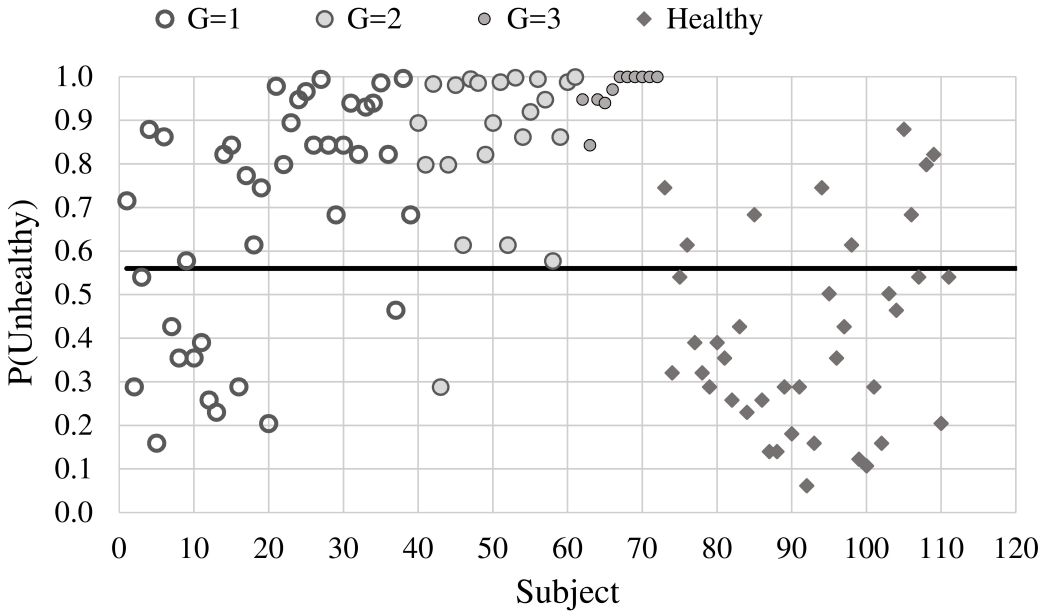

Fig. 6.3 Fitted values of the best logistic regression model for the reading task acquired with the headworn microphone, in terms of probability of having unhealthy voice. Circle points indicate the patient group, where different colours and sizes represent subjects with different overall grade of dysphonia; diamond points indicate the control group. The bold line represents the threshold value of 0.56, which best separates patients and control subjects.

voice disorders and people who reveal their vocal problems while speaking. The best classification threshold was P(Unhealthy) = 0.56, that corresponds to 18.0 dB in terms of $CPPS_{95\text{prc}}$. Note that several patients who are wrongly classified as healthy subjects were labelled with G equal to 1, which indicates the lowest dysphonia rate.

Table 6.3 summarizes the results obtained for the speech tasks acquired with the ECM. The *p*-values of the Two-tailed Mann-Whitney U-test highlight that all the CPPS parameters are significantly different in healthy and pathological voices, with some exceptions, for both the speech materials. However, no model with a CPPS parameter as independent variable and the presence/absence of voice disorders as dependent variable shows a sufficient performance on classifying healthy and unhealthy voice: all the models have an AUC lower than 0.80, thus not having a good discriminator power. Then, no model has been selected as the best in discriminating between healthy and pathological voices in the case of the ECM. As a general comment, it is noticeable that the models related to reading have higher AUC than

Table 6.3 The same of Table 6.1. Data refers to reading and free speech recorded with the ECM.

CPPS param.	**Reading**				**Free speech**			
	U-test	Mc Fad.	AUC (CI)	Acc.	U-test	Mc Fad.	AUC (CI)	Acc.
$CPPS_{\text{mean}}$	**0.001**	0.17	0.77 (0.68-0.86)	69%	**0.001**	0.14	0.74 (0.65-0.83)	64%
$CPPS_{\text{median}}$	**0.001**	0.19	0.78 (0.68-0.86)	69%	**0.001**	0.15	0.74 (0.64-0.83)	65%
$CPPS_{\text{mode}}$	**0.001**	0.20	0.77 (0.69-0.86)	70%	**0.001**	0.20	0.75 (0.66-0.84)	65%
$CPPS_{\text{std}}$	**0.001**	0.11	0.71 (0.61-0.81)	67%	**0.001**	0.11	0.71 (0.61-0.81)	69%
$CPPS_{\text{range}}$	**0.001**	0.18	0.77 (0.68-0.86)	65%	**0.001**	0.22	0.79 (0.71-0.88)	72%
$CPPS_{5\text{prc}}$	**0.009**	0.01	0.65 (0.54-0.76)	58%	0.091	0.01	0.60 (0.48-0.72)	59%
$CPPS_{95\text{prc}}$	**0.001**	0.15	0.76 (0.68-0.85)	69%	**0.001**	0.17	0.76 (0.67-0.85)	64%
$CPPS_{\text{skew}}$	**0.001**	0.13	0.75 (0.66-0.84)	67%	**0.001**	0.11	0.70 (0.60-0.80)	61%
$CPPS_{\text{kurt}}$	**0.006**	0.01	0.66 (0.56-0.77)	57%	0.090	0.01	0.60 (0.49-0.71)	60%

those related to free speech and AUCs of reading are more comparable than those of free speech.

Within-speaker consistency across tasks

In order to investigate the consistency across tasks of the investigated measures, Pearson coefficient between the values of each CPPS parameter from reading and free speech has been determined. Table 6.4 shows that all the CPPS parameters are significantly correlated between reading and free speech acquired with the headworn microphone, where the most highly correlated are $CPPS_{\text{mean}}$, $CPPS_{\text{median}}$ and $CPPS_{95\text{prc}}$, with the highest correlation coefficient of 0.94 for the latter parameter.

The CPPS parameter included in the best model for discriminating healthy and pathological voices shows also the best consistency across tasks. Once more, such results obtained from the descriptive statistics for CPPS distributions highlight that the within-speaker consistency across tasks is also valid for CPPS distributions themselves. Figure 6.4 shows overlapped CPPS distributions from reading and free speech acquired with the headworn microphone from 20 healthy subjects: each subject obtained a couple of CPPS distributions with very similar shape and central tendency. The same observation can be done in Figure 6.5, which includes overlapped CPPS distributions from reading and free speech from 20 patiences. Such permanent characteristic has been labelled as "CPPS *vocal*print" by my research group: just like a *finger*print, it is "personal", as a subject can be recognized by means of its CPPS distribution, and "permanent", as it describes the subject's voice status using whatever speech material.

Table 6.4 Pearson coefficients between CPPS values obtained from reading and free speech, for the two microphones.

CPPS parameter	**Pearson coeff.**	
	Headworn microphone	ECM
$CPPS_{\text{mean}}$	0.92	0.76
$CPPS_{\text{median}}$	0.92	0.77
$CPPS_{\text{mode}}$	0.81	0.77
$CPPS_{\text{std}}$	0.82	0.60
$CPPS_{\text{range}}$	0.78	0.73
$CPPS_{\text{5prc}}$	0.73	0.71
$CPPS_{\text{95prc}}$	*0.94*	*0.81*
$CPPS_{\text{skew}}$	0.88	0.66
$CPPS_{\text{kurt}}$	0.71	0.62

Table 6.4 also shows that CPPS parameters obtained from reading and free speech using the ECM have lower correlation coefficients than the ones of the headworn microphone. However, $CPPS_{\text{95prc}}$ shows the highest consistency across tasks again, with a Pearson coefficient of 0.81. Figures 6.6 and 6.7 show overlapped CPPS distributions from reading and free speech acquired using the ECM from 20 control speakers and 20 patiences, respectively: from a qualitative point of view the "CPPS vocalprint" is less recognizable in each subject, since more CPPS variations in the two speech tasks are present.

Consistency of measures across tasks with various phonemic content is a qualification that proves the clinical and research utility of the parameters. In the present work, several CPPS parameters resulted moderately to strongly correlated between reading and free speech tasks, therefore individual CPPS distributions were similar across the two speech materials, as shown in Figures 6.4, 6.5, 6.6 and 6.7. These outcomes extend the strong relationship (correlation coefficient of 0.962) that Lowell *et al.* [81] found between CPPS mean in a long sentence and in a short constituent phrase acquired with a microphone in air, for the dysphonic and normal speakers. Due to the strong correlations between CPPS parameters in reading and free speech acquired with the headworn microphone that have been found in this work, the best empirical logistic models for reading and free speech were comparable and with similar threshold values. The sensitivity and specificity of the threshold value for the

reading task are higher than those of the free speech-model thanks to the identical phonemic content that reading a passage guarantees for all the speakers.

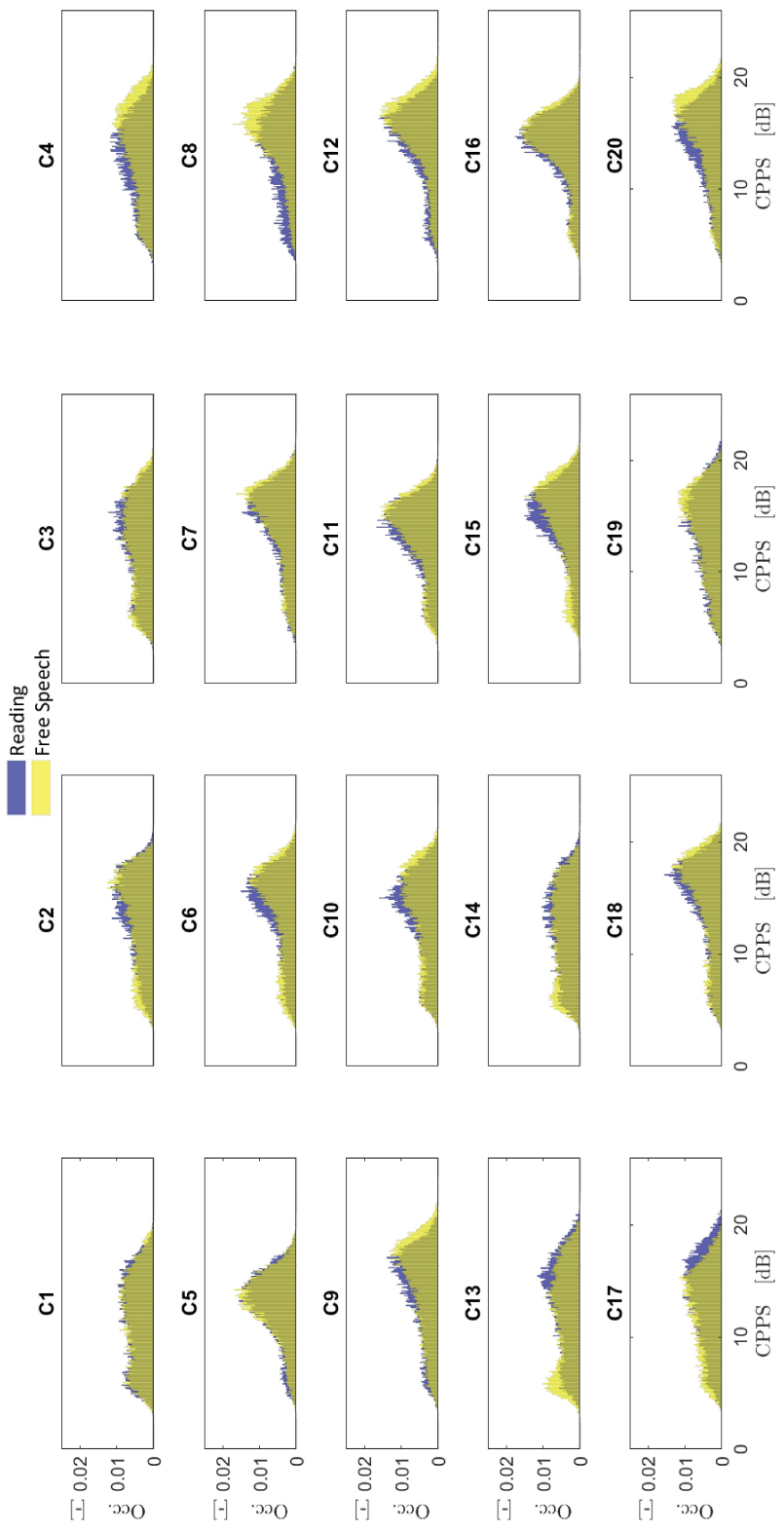

Fig. 6.4 CPPS distributions obtained from the analyses of reading (blue) and free speech (yellow) tasks acquired with the headworn microphone from 20 healthy subjects.

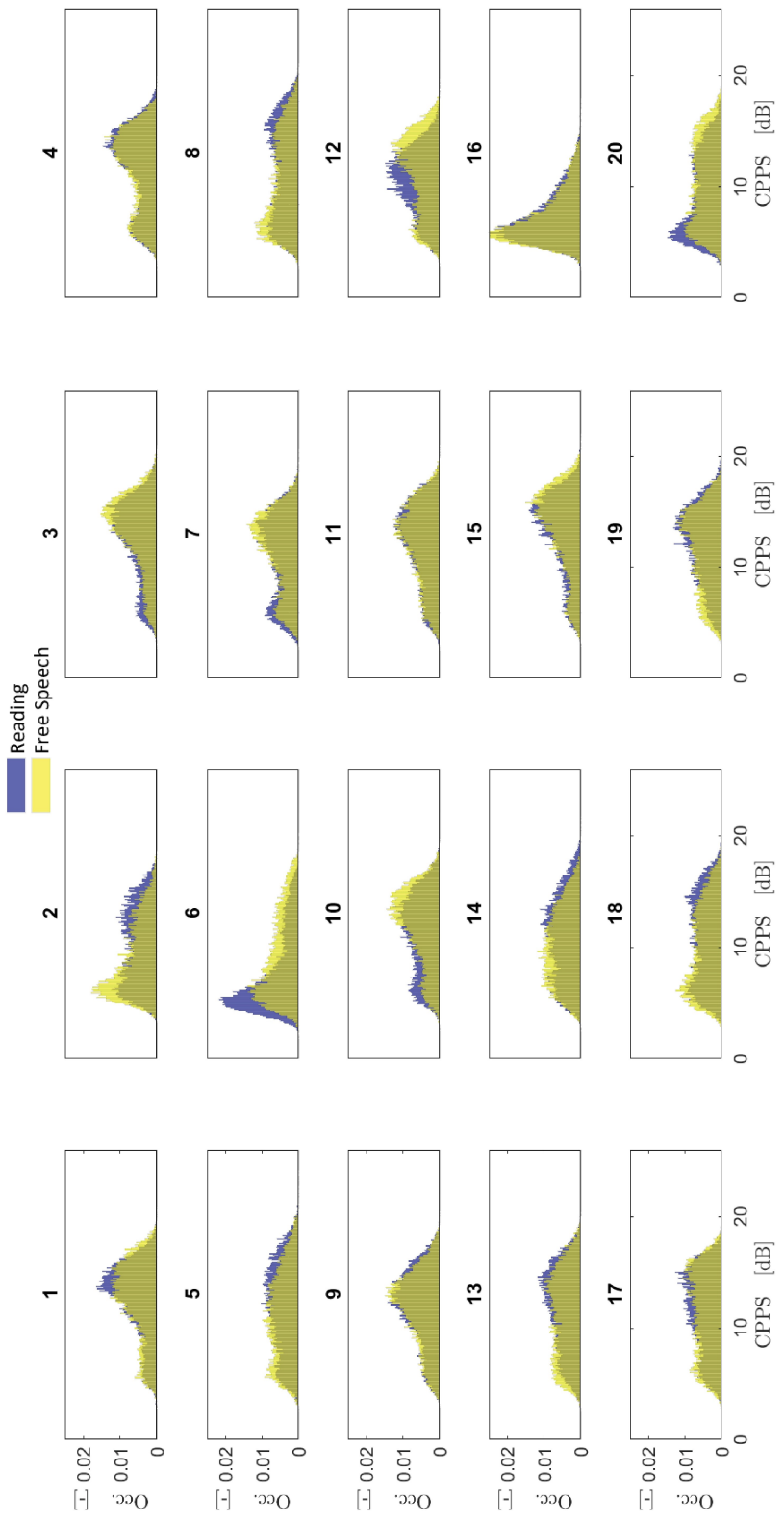

Fig. 6.5 CPPS distributions obtained from the analyses of reading (blue) and free speech (yellow) tasks acquired with the headworn microphone from 20 patients.

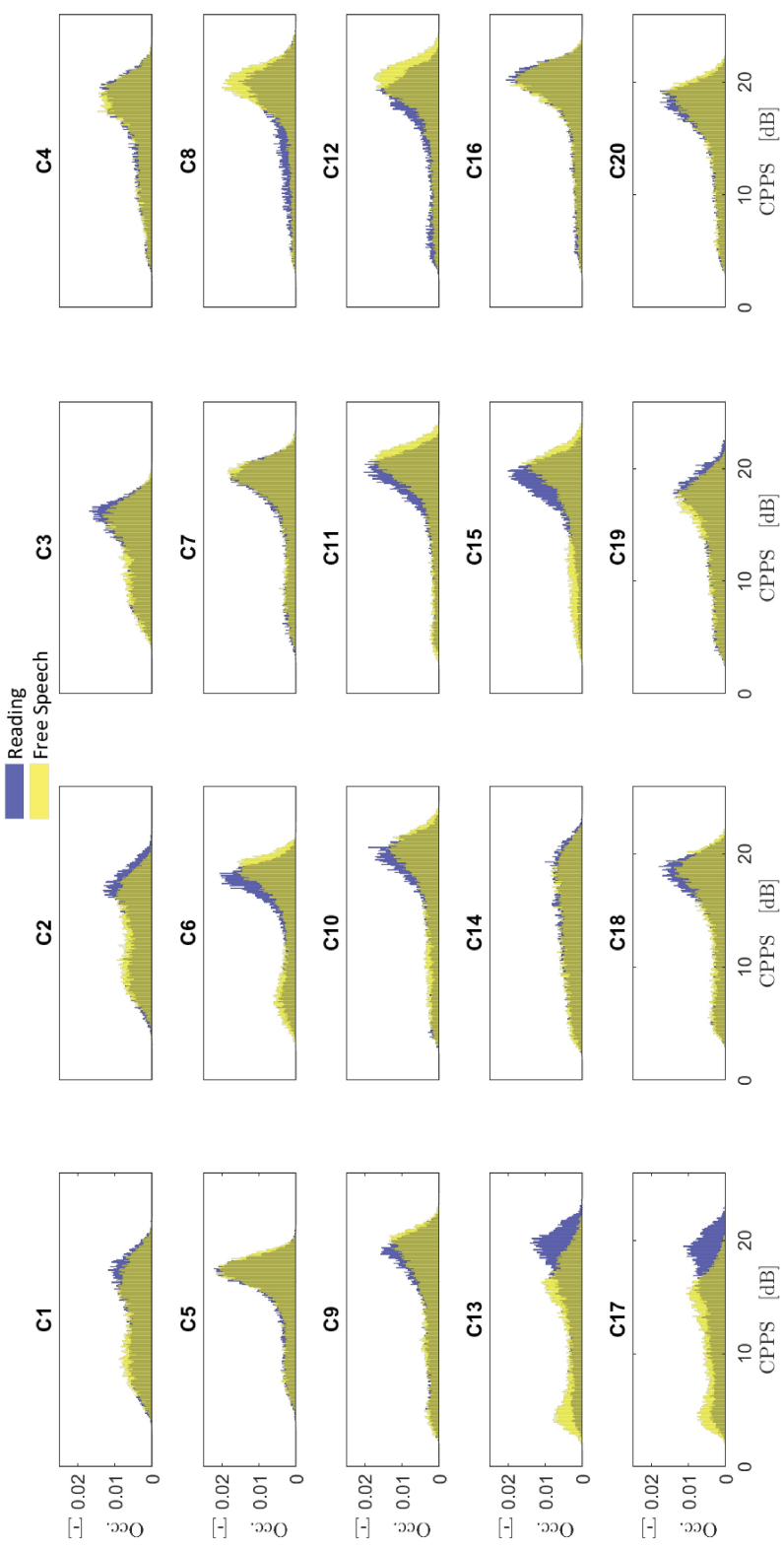

Fig. 6.6 CPPS distributions obtained from the analyses of reading (blue) and free speech (yellow) tasks acquired with the ECM from 20 healthy subjects.

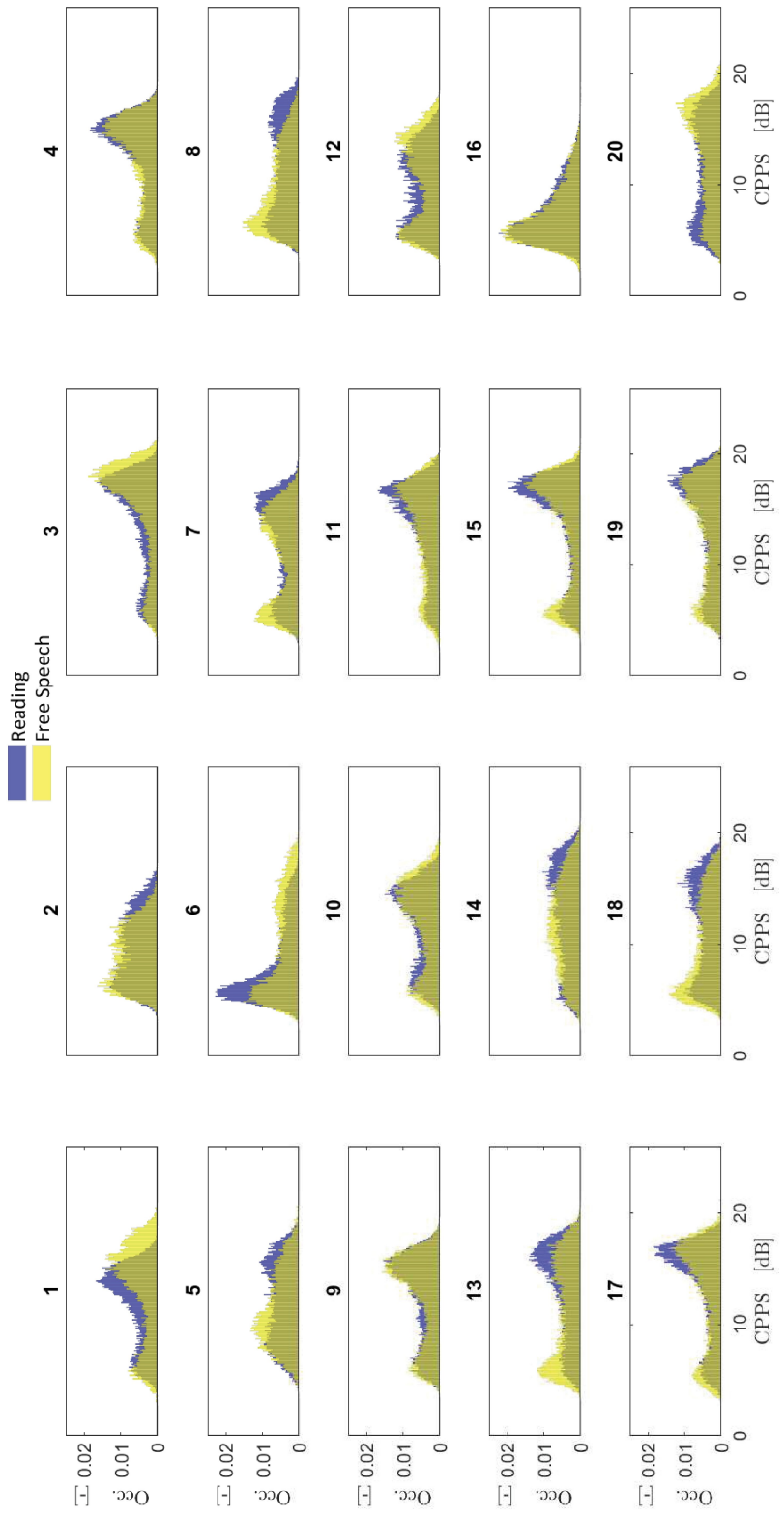

Fig. 6.7 CPPS distributions obtained from the analyses of reading (blue) and free speech (yellow) tasks acquired with the ECM from 20 patients.

Monte Carlo method

Due to the strong similarity between the two logistic models obtained for reading and free speech acquired with the headworn microphone, the uncertainty estimation of the threshold value has been performed including CPPS parameters from both the speech materials for each subject, thus accounting for both the intra- and inter-speaker variability. The best-fitted distributions of the parameter $CPPS_{95prc}$ for pathological voices is the bimodal, while for healthy speakers is the Weibull one. Their probability density functions have been used for the implementation of the Monte Carlo method based on 1000 trials, as described in paragraph 5.3.4. The obtained distribution of threshold-values has a 95% confidence interval equal to 0.6 dB, which constitutes the $CPPS_{95prc}$ threshold variability for continuous speech. This interval is represented as a gray area around the threshold in Figure 6.8, where $CPPS_{95prc}$ values for the readings acquired with the headworn microphone are represented.

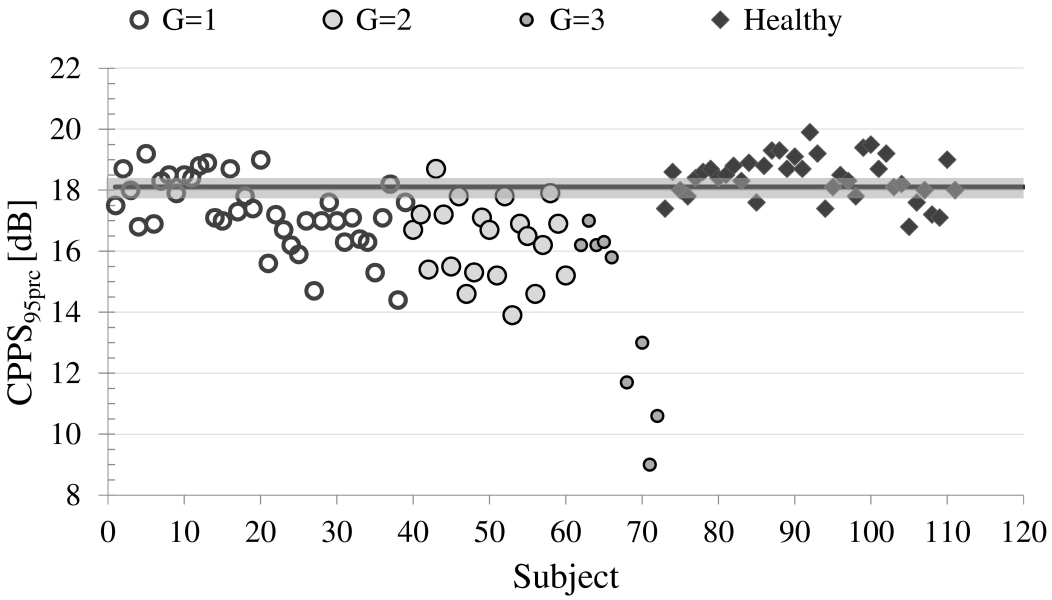

Fig. 6.8 $CPPS_{95prc}$ values for the readings acquired with the headworn microphone. Circle points indicate the patient group, where different colours and sizes represent subjects with different overall grade of dysphonia; diamond points indicate the control group. The bold line indicates the threshold value (18.1 dB) and the gray area corresponds to its 95% confidence interval.

Voice self-assessment

As described in paragraph 6.2.1, participants filled the Italian version of the Voice Activity And Participation Profile [70], i.e., the Profilo di Attività e Partecipazione vocale, PAPV, [187]. The PAPV is a 28-item questionnaire that investigates voice activity limitation and participation restriction in pathological subjects. The items are rated on a 10 cm-visual analogue scale and are divided into five sections:

1. self-perceived voice problem (severity): 1 question, maximum score 10;
2. job section (job): 4 questions, maximum score 40;
3. daily communication (daily): 12 questions, maximum score 120;
4. social communication (social): 4 questions, maximum score 40;
5. emotional section (emotional): 7 questions, maximum score 70;

The minimum possible total score is zero, while the maximum is 280.

Sixty-eight patients and thirty-eight controls were selected, since they performed both the reading and free speech tasks and filled the PAPV. Among them, 44 patients and 28 controls were workers: the investigation on voice self-assessment for this group of subjects has been performed using the whole questionnaire. The job section has been discarded for the rest of participants, thus considering a total score equal to 240.

Table 6.5 Average PAPV value, *Mean*, and relative standard deviation, *SD*, for pathological and healthy voices. The section *Job* is discarded because both workers and non-workers are included.

PAPV	**Pathological voices**		**Healthy voices**	
	Mean	SD	Mean	SD
Severity	5.3	3.2	0.7	1.2
Job	-	-	-	-
Daily	37.0	29.5	10.8	32.7
Social	7.2	8.7	0.9	3.0
Emotional	21.5	16.3	3.7	8.2
Total	71.5	50.0	13.3	23.1

Table 6.6 Average PAPV value, *Mean*, and relative standard deviation, *SD*, for pathological and healthy voices. Only the workers are included.

PAPV	Pathological voices		Healthy voices	
	Mean	SD	Mean	SD
Severity	5.7	2.9	0.9	1.4
Job	13.6	10.2	2.8	5.9
Daily	38.3	24.5	13.9	37.7
Social	7.1	8.4	1.0	3.5
Emotional	21.8	16.0	4.3	9.5
Total	86.5	49.7	16.3	31.7

Tables 6.5 and 6.6 show PAPV scores obtained from the patients and the healthy subjects, when considering all the participants and only the workers, respectively. The two tables indicate that significantly higher values have been found in each PAPV section for pathological voices with respect to the controls. The average of PAPV scores of each section are comparable to the results reported by Fava *et al.* [187], where 239 Italian individuals (108 with vocal disorders and 131 without voice problems) filled the same questionnaire.

Figure 6.9 shows the rating proportions of PAPV sections for the pathological workers. The highest percentage of ratings (33%) comes from the section of self-perceived voice problem; the sections related to job, daily communication and emotions have similar percentages equal to 20%, 19% and 18%, respectively; the social communication section has the lowest ratings (10%). Similar results have been found for all the pathological subjects, when the job section has been discarded for the analyses: figure 6.10 shows the same evidences, with higher percentages for each section. These outcomes indicate that patients are aware of their vocal problem and that their quality of life is affected by participation restriction and activity limitation in different life-related aspects.

As a further investigation, the strength of relationships between the best CPPS parameter in discriminating between healthy and pathological voice in continuous speech (paragraph 6.2), $CPPS_{95\text{prc}}$, and PAPV scores in both the groups of speakers have been examined. The correlations analyses have been performed using the Spearman coefficient, due to the presence of discrete data as PAPV ratings. Table 6.7 summarizes the correlations between $CPPS_{95\text{prc}}$ obtained from reading and free speech and PAPV scores, when all the participants or only the workers are considered:

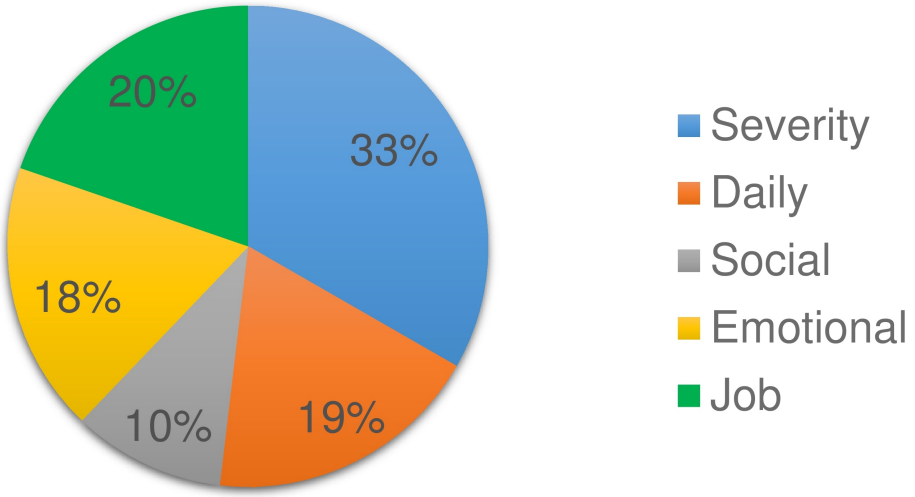

Fig. 6.9 Percentages on average score obtained for each PAPV section over the workers with dysphonia.

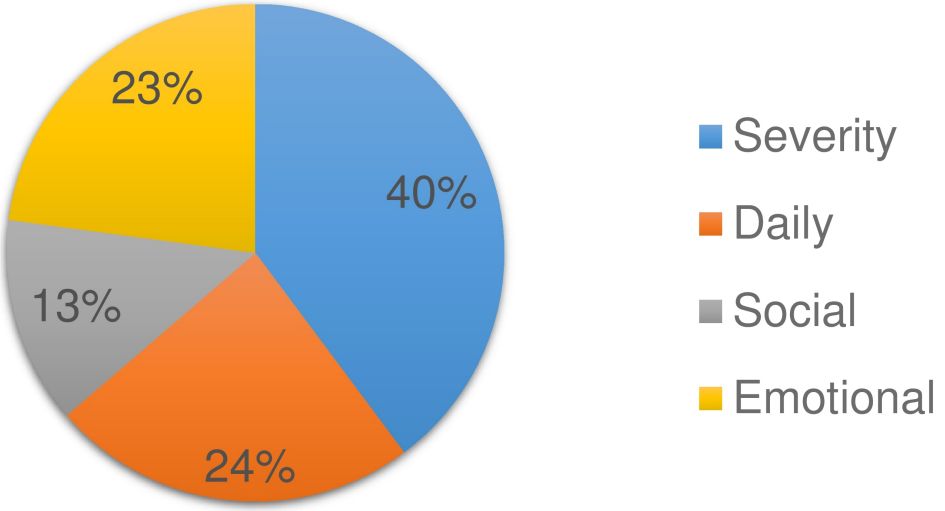

Fig. 6.10 Percentages on average score obtained for each PAPV section over all the subjects with dysphonia.

all the coefficients indicate moderate strength correlations in all the PAPV sections for both the speech materials. The negative sign of the coefficients designates an expected inverse correlation, since $CPPS_{95\text{prc}}$ decreases with the presence of vocal disorders. The PAPV total score shows the highest correlations of about -0.60 in all the cases, while generally the emotional section has the lowest coefficients in both

reading and free speech. The sections that show the best correlation between PAPV ratings and $CPPS_{95prc}$ are the ones related to daily communication and self-perceived voice problem. Slight differences can be observed between the two speech materials: $CPPS_{95prc}$ from reading is more correlated with the scores of self-perceived voice problem and emotional sections, while $CPPS_{95prc}$ from free speech better correlates with daily communication ratings. These results confirm and extend the study conducted by Awan *et al.* [188], where a Spearman coefficient of -0.50 was found between CPP mean from readings and the Voice Handicap Index scores, in 258 patients and 74 control subjects. As already highlighted in this paper, although both cepstral analyses and self-report methodologies may be sensitive metrics of voice impairment and disability, note that they provide meaningful and complementary information that allows to a multidimensional assessment of voice.

Table 6.7 Spearman correlation coefficients for $CPPS_{95prc}$ obtained from reading and free speech versus PAPV scores (all p-values < 0.001).

PAPV	**All subjects**		**Workers**	
	Reading	Free speech	Reading	Free speech
Severity	-0.53	-0.45	-0.57	-0.48
Job	-	-	-0.56	-0.52
Daily	-0.55	-0.57	-0.54	-0.58
Social	-0.51	-0.54	-0.51	-0.55
Emotional	-0.50	-0.48	-0.48	-0.42
Total	-0.61	-0.60	-0.60	-0.59

6.3 Study 2: Cepstral Peak Prominence Smoothed distribution in continuous speech with different voice qualities

These studies continue the investigation on clinical applications for CPPS distribution in continuous speech. Changes in CPPS distributions in various voice qualities using different types of microphones have been determined in paragraph 6.3.1. CPPS distribution may help in the diagnostic procedure and furnish proof of outcomes after interventions such as voice therapy and phonosurgery. Paragraph 6.3.3 presents preliminary results in such clinical applications.

6.3.1 First experiment

In a sound-treated booth, 5 voice experts (2 females and 3 males) read a Swedish text of 88 words producing 4 different voice qualities, namely "Normal", "Creaky", "Breathy" and "Strain" voice. The text, which is reported in the appendix A, is used routinely as speech material in the standardized recording setting at the Department of Speech Language Pathology at the Karolinska University Hospital. Each participant worn the following 3 devices while performing the task (Figure 6.11).

1. A headset microphone (Sennheiser MKE-2, USA), MIC, which was mounted at 15 cm distance from the mouth. It was connected to a PC audio board, which samples the signal at a rate of 16 kSa/s and 16 bit of resolution.

2. An electret condenser microphone, ECM, (AE38 Alan Electronics GmbH, Dreieich, Germany), which was fixed at the jugular notch by means of a surgical band. It was connected to the handy recorder ROLAND R05 (Roland Corp., Milano, Italy) that samples the signal at a rate of 44.1 kSa/s using 16 bit of resolution.

3. A piezoelectric contact microphone, PM, (HX-505-1-1, Shenzhen, China), which is embedded in a collar placed around the neck and connected through an AUX cable to a smartphone (Samsung SM-G310HN). The recordings were performed using the Vocal Holter App (PR.O.VOICE, Turin, Italy) and saved into the internal memory of the smartphone using a rate of 22.05 kSa/s and 16 bit of resolution.

Figure 6.12 shows the three devices: differently from the MIC that acquires voice signals at the output of the lips, the ECM and PM are two contact microphones that sense the skin vibrations induced by the vocal-fold activity.

6.3.2 Analyses and Results

Firstly, all the acquired signals have been re-sampled to 22050 Hz and the signals from the three devices have been aligned for each reading. A CPPS distribution has been obtained for each reading, according to the computation described in 6.1. Figure 6.13 shows overlapped CPPS distributions for "Normal", "Creaky", "Breathy"

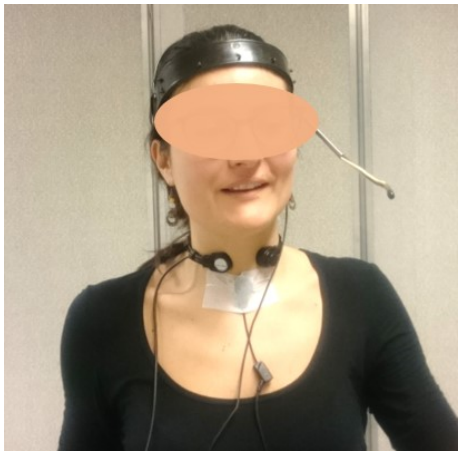

Fig. 6.11 A participant while performing the experiment with the three devices.

Fig. 6.12 The 3 devices used for acquiring the voice signal: a) the headset microphone (MIC); b) the piezoelectric microphone (PM) and c) the condenser microphone (ECM).

and "Strain" voice qualities, obtained from a female and a male subject with the three devices.

With the aim of investigating the main reasons of the differences among CPPS distributions of different voice qualities acquired from different devices, the frequency response of each acquisition chain has been deepened. Figure 6.14 shows the Long Term Average Spectrum (LTAS) of "Normal" voice readings after removing unvoiced segments for each measurement chain. A first comment concerns the LTAS slope. Since the signal acquired at the output of the contact sensors is affected by the physiological filtering (vocal folds – throat – skin), but not by the filtering effect of the vocal tract which instead affects microphones in air, a high slope for the two contact sensors was expected. However, the LTAS of PM has a boost of energy

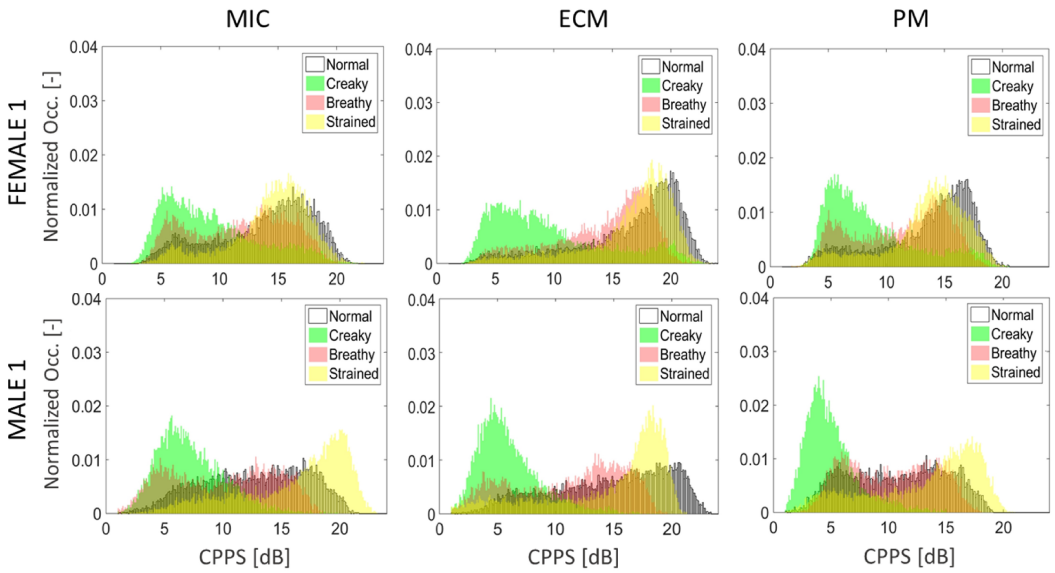

Fig. 6.13 CPPS distributions for different voice qualities, obtained from a female and a male subject with the three devices.

content between 2 kHz and 4 kHz due to the usual use of them: they are used in very noisy environments, such as in the chockpit by helicopter drivers, then such energy boost helps intelligibility. Figure 6.14 also underlines the frequency content in the acquired signals: the LTAS of MIC reaches the maximum energy content of 7.5 kHz, since the sampling rate of the acquisition system was of 16000 Hz; for the two contact sensors, instead, a lower frequency content is noticeable: 3.5 kHz for the ECM and 5 kHz for the PM. Furthermore, Figure 6.14 shows the noise level in the signals acquired with the three devices: the noise level is 60 dB lower than the peak magnitude in both LTAS from MIC and ECM, while it is 50 dB lower the peak for PM, thus meaning that a higher noise content is acquired with the PM chain.

In summary, readings acquired with ECM have a limited frequency content, while signals acquired with PM have a higher noise level in the spectrum. Figure 6.15 confirms for the reading the main consequences in CPPS values already discussed in Chapter 5 for the vowel /a/: such limited frequency content provides higher CPPS values for ECM, i.e. CPPS distributions from ECM are shifted to the right with respect to the CPPS distributions from MIC; moreover, the higher noise level in signals from PM leads to lower CPPS values, i.e. CPPS distributions from PM are shifted to the left with respect to the CPPS distributions from MIC and ECM.

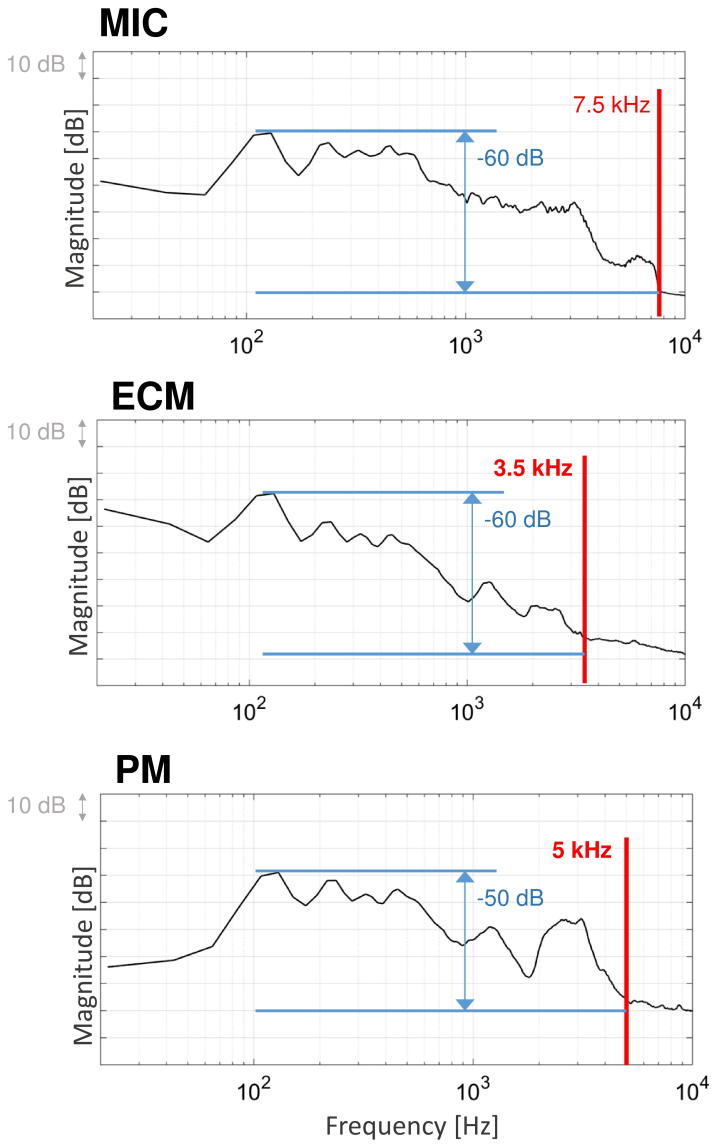

Fig. 6.14 Long Term Average Spectra (LTAS) of a "Normal" voice for each device.

Actually, in Figure 6.15 there is a distribution that does not follow the before-mentioned shifting, that is the one related to the "Strained" voice: such CPPS distribution has higher values than the "Normal" one for both MIC and PM, while for EMC it is shifted to the left with respect to the "Normal" distribution. Figure 6.16 shows the overlapped LTAS of the different voice qualities for each device: "Strained" voice acquired with MIC and PM has higher CPPS values, since such voice quality does not introduce any irregularity in the spectrum, but higher peaks

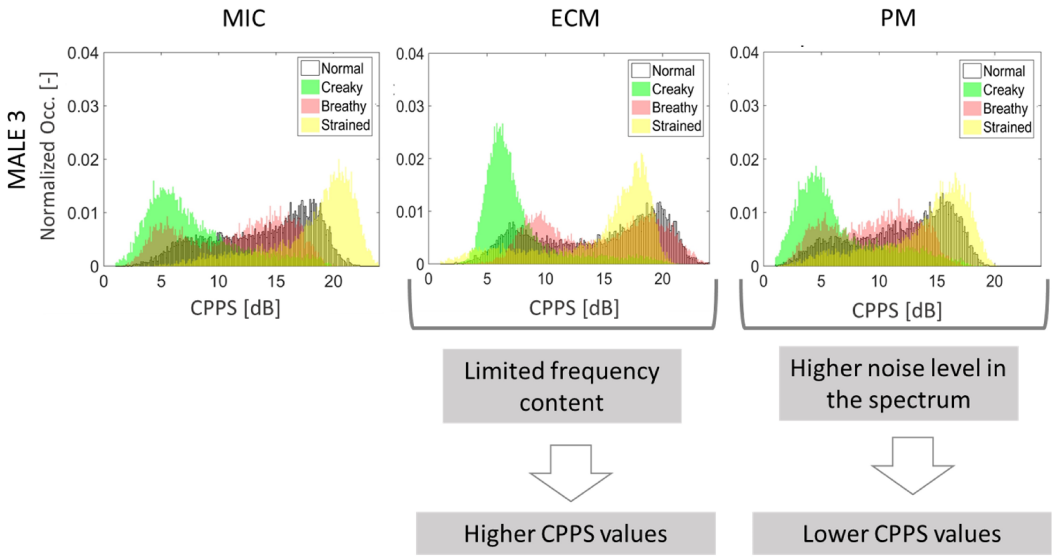

Fig. 6.15 Relationships between CPPS distributions and spectral characteristics.

in harmonics are present thus magnifying the spectrum periodicity up to higher frequencies with respect to "Normal" voice quality [74]; differently, ECM cannot peak up harmonic peaks with high energy content in the case of "Strained" voice, so that a low number of spectral harmonics are computed by the CPPS algorithm and CPPS values drop.

CPPS distributions in figures 6.13 and 6.15 highlight different characteristics both in shape and central tendency for each voice quality. Figure 6.17 corroborates such a note, since it shows overall CPPS distributions from all the participants for each device: from a qualitative point of view, CPPS distributions from the three devices are are similar and they have similar marks for the same voice quality. As already highlighted in 6.2, CPPS distributions of "Normal" voice quality have most occurrences in high CPPS values that proves the regularity in the spectrum for the mayority of the continuous speech material. However, low CPPS values are also present because of few speech sounds in continuous speech that produce irregular spectra (e.g. consonants or fundamental frequency changing). CPPS distributions from "Strained" voice have a similar shape of the "Normal" voice, due to the before-mentioned causes, while CPPS distributions from "Breathy" and "Creaky" voice show highly different shapes. "Breathy" voice has a CPPS distribution that is shifted on the left with respect to the "Normal" CPPS distribution, since "Breathy" voice

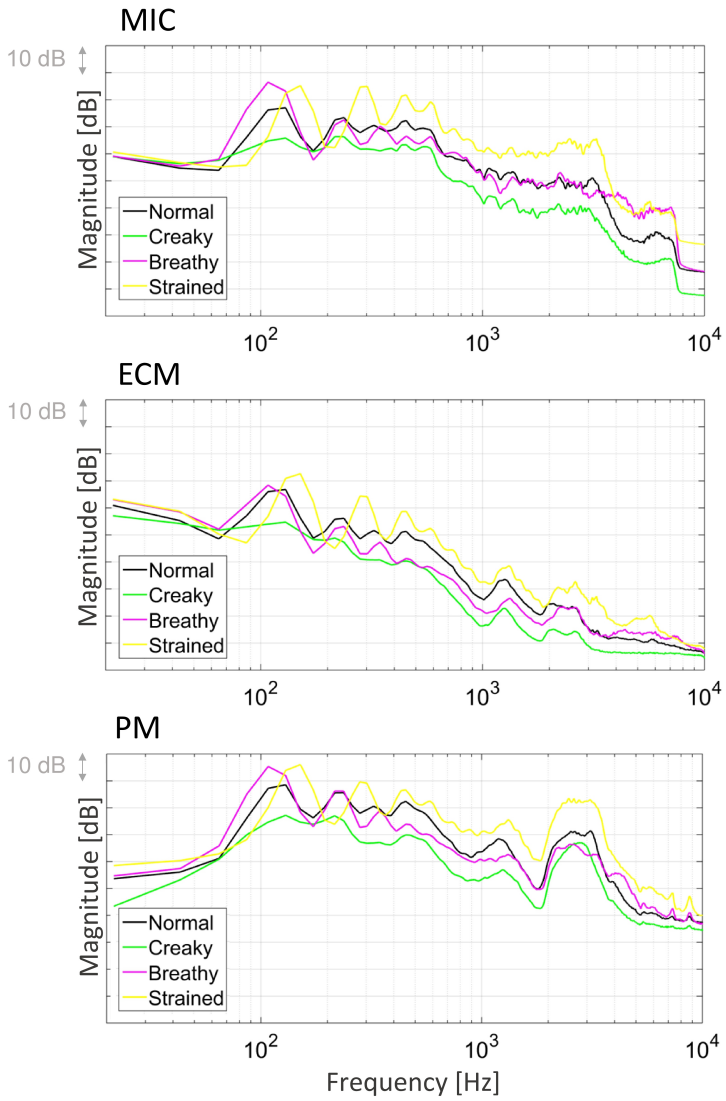

Fig. 6.16 Overlapped Long Term Average Spectra (LTAS) of different voice qualities for each device.

is characterized by audible air escape during voice production that produces noise in high frequences of the spectrum. Moreover, CPPS distribution from a "Breathy" voice is a bimodal distribution, where both the modes have similar occurrences, because of a common behaviour among the participants to the experiment: they introduced an increasing "breathiness effect" in the conclusive syllables of the words, thus producing half occurrences of very low CPPS values. "Creaky" voice shows a CPPS distribution that, contrary to the "Normal" voice quality, has most occurrences

Table 6.8 Pearson correlation coefficients between CPPS parameters obtained from signals acquired with the three devices (* p-value<0.05; ** p-value<0.01; *** p-value<0.001).

CPPS parameter	Correlation coeff. MIC-ECM	Correlation coeff. MIC-PM	Correlation coeff. ECM-PM
$CPPS_{\text{mean}}$	0.87***	0.93***	0.94***
$CPPS_{\text{median}}$	0.92***	0.95***	0.97***
$CPPS_{\text{mode}}$	0.92***	0.95***	0.97***
$CPPS_{\text{std}}$	0.90***	0.87***	0.89***
$CPPS_{\text{range}}$	0.54*	0.78***	0.74***
$CPPS_{5\text{prc}}$	0.46*	0.60**	0.54*
$CPPS_{95\text{prc}}$	0.83***	0.91***	0.89***
$CPPS_{\text{skew}}$	0.96***	0.97***	0.97***
$CPPS_{\text{kurt}}$	0.82***	0.80***	0.76***

in low CPPS values. Such result is due to the continuous irregularity in the spectrum obtained by very shortened and slackened vocal folds that produce such a voice quality [74].

Table 6.9 shows Pearson coefficients between paired CPPS parameters obtained from signals acquired with the three devices and strong correlations result for all the parameters (Pearson coefficient higher than 0.80), except for $CPPS_{\text{range}}$ and $CPPS_{5\text{prc}}$ that reach the lowest coefficient of 0.50 between MIC and ECM. Such outcomes confirm the previous qualitative observations on CPPS distributions obtained from the different microphones, thus highlighting that both their central tendency and their overall shape is kept when the same type of voice quality is analysed.

6.3.3 Second experiment

In a sound treated booth, 13 Swedish patients of the Department of Clinical Science, Intervention and Technology (CLINTEC) at Karolinska Institutet in Stockholm read the Swedish text while wearing the headset microphone as described in 6.4. In particular, 6 patients read the text twice: 5 subjects before starting the speech therapy and at the end of the period and 1 patient performed the experiment before and after the surgical intervention for removing a vocal fold edema.

Two expert speech pathologists of the department performed the audio-perceptual assessment of voice of the 20 collected readings in consensous: they individually

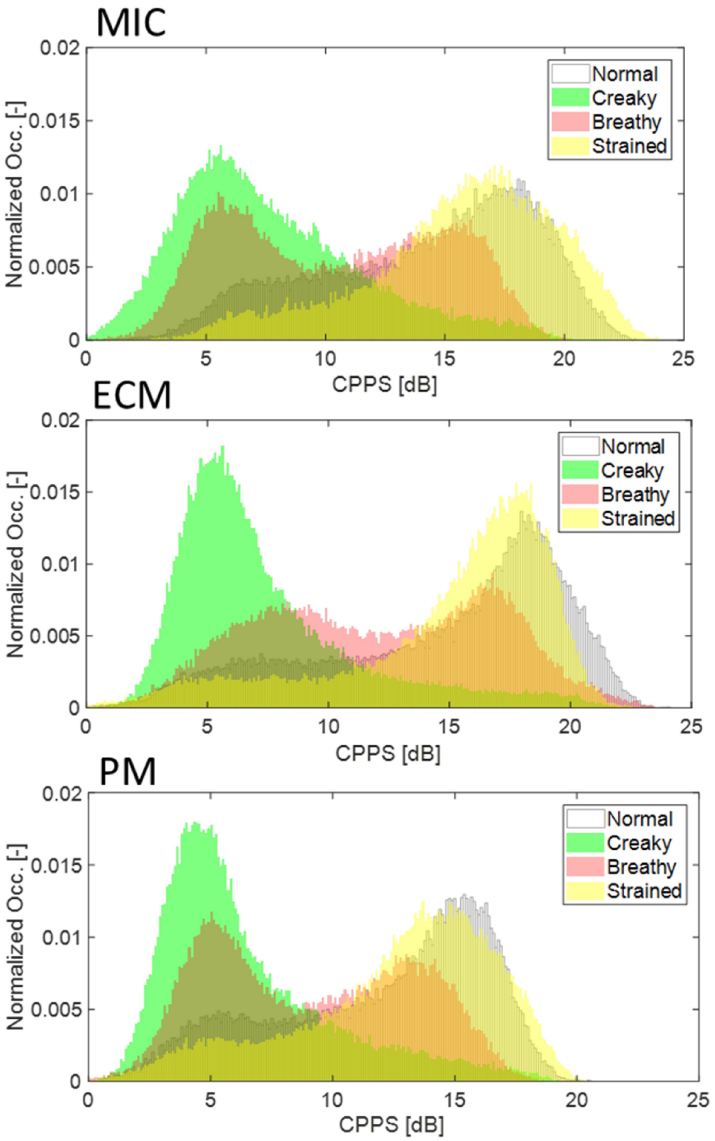

Fig. 6.17 Overall CPPS distributions from all the participants for each voice quality and device.

rated each reading using the perceptual voice evaluation sheet showed in Figure 6.18, which follows the Stockholm Voice Evaluation Approach (SVEA)[72], and then they discussed on each rate until they were totally in agreement. The perceptual voice evaluation sheet presents a visual analogue scale, that is a 10-cm line scale with unlabelled anchors, to assess each of the voice quality features. The left extremity of the line reflects the absence of the quality being judged, while the right end

of the line reflects the listener's judgement of the most deviant example for the voice quality feature under judgement. A tick mark is drawn on the line to reflect a listener's judgement for each voice quality and the length from the left end to each tick mark quantifies the presence of each voice quality feature in the voice sample under analysis. The two expert speech pathologists also rated in consensus 36 Italian readings acquired with the headworn microphone, which were performed by 36 subjects from the patient group as described in 6.2.

6.3.4 Analyses and Results

CPPS distributions and the respective descriptive statistics have been computed for each reading, using a sample frequency of 22050 Hz and a window size of 1024 samples, as in the previous investigations. Figures 6.19 and 6.20 show the auditory perceptual evaluation of voice and CPPS distributions of 5 patients who were recorded before and after the speech therapy period and of the patient who was recorded before and after the surgery.

These figures allow observing which aspects of voice quality are more represented than others by CPPS distribution. The first subject in Figure 6.19, as example, had an high rate of breathiness and a moderate rate of aphonia that resulted in a limited rate of breathiness after the speech therapy; CPPS distribution completely changes its shape and central tendency in the two stages, thus highlighting that it reflects the two prevalent voice qualities. The second subject in Figure 6.19 and the third subject in Figure 6.20, instead, had a predominant rate of instability at the beginning that after the therapy disappeared or began low: however, such a great improvement in voice quality has not been mirrored in CPPS distributions, since they kept the shape and slightly shifted on the right in the second stage. Consequently, as a preliminary analysis, it can be stated that the change in CPPS distribution observed for the subject 4 in Figure 6.20 is mostly due to the deleting of breathiness in his voice quality and not to the instability solving. Moreover, subject 3 in Figure 6.19 and subject 5 in Figure 6.20 respectively show that creakyness and roughness are well assessed by CPPS distribution. Table 6.9 summarizes all the correlation coefficients calculated between the perceptual assessment of voice and CPPS distribution obtained for the readings from Sweden and Italian patients. Only those voice quality features that were present in at least 15 subjects with heterogeneous ratings has been included

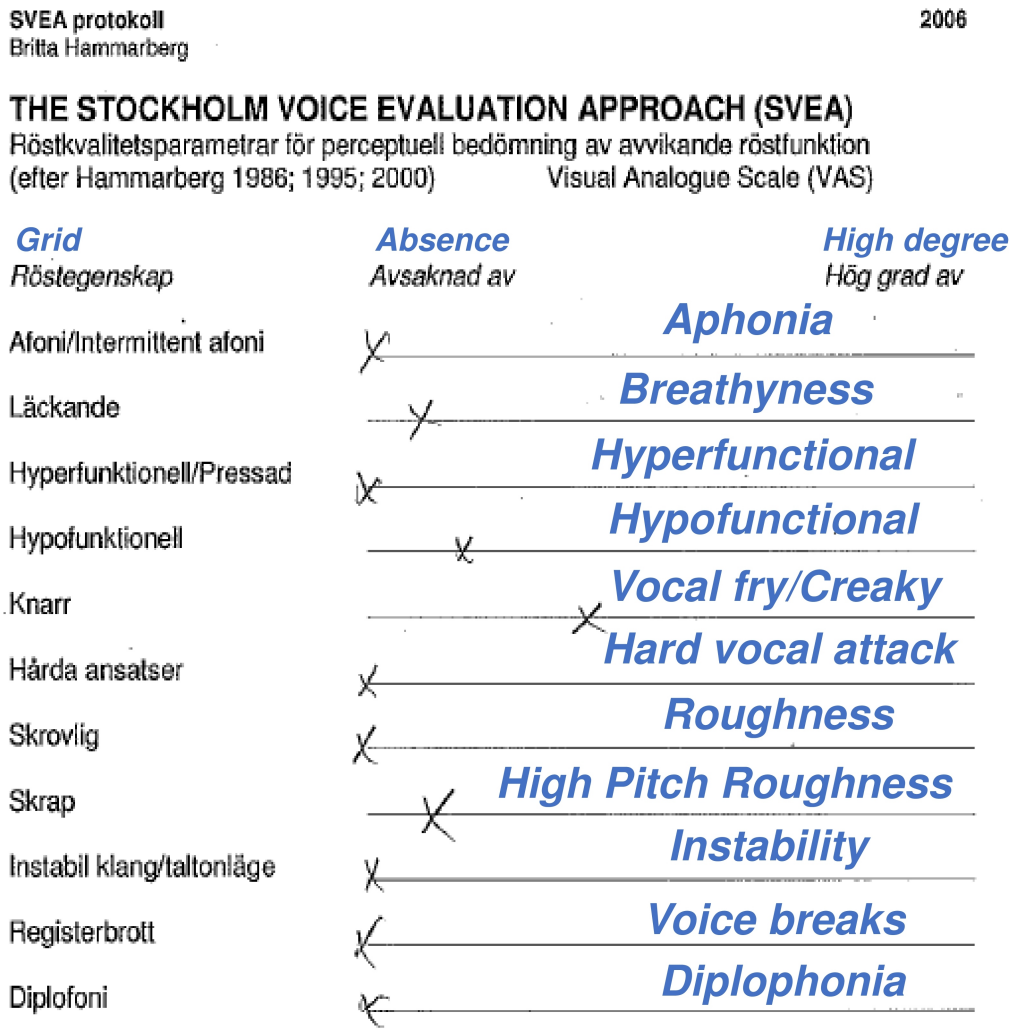

Fig. 6.18 The voice evaluation sheet.

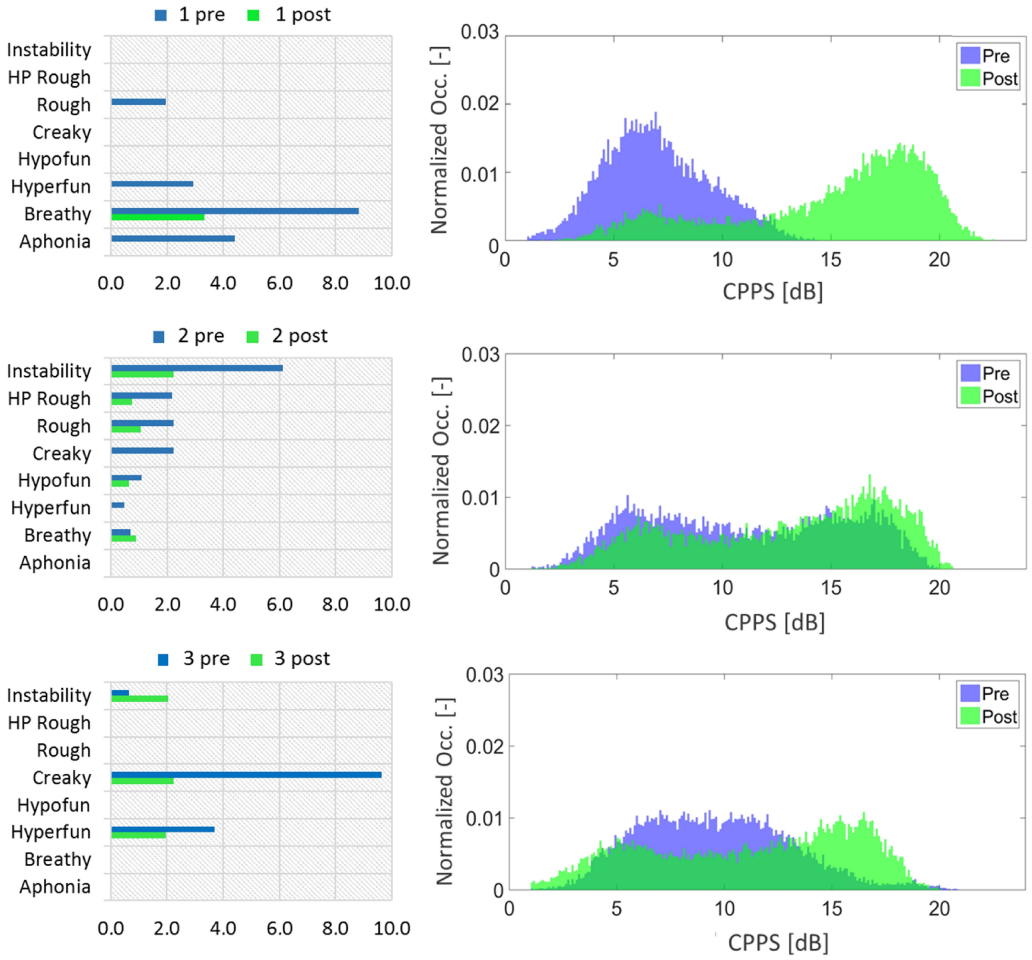

Fig. 6.19 Auditory perceptual evaluation, on the left, and CPPS distributions, on the right, before (blue) and after (green) the speech therapy period.

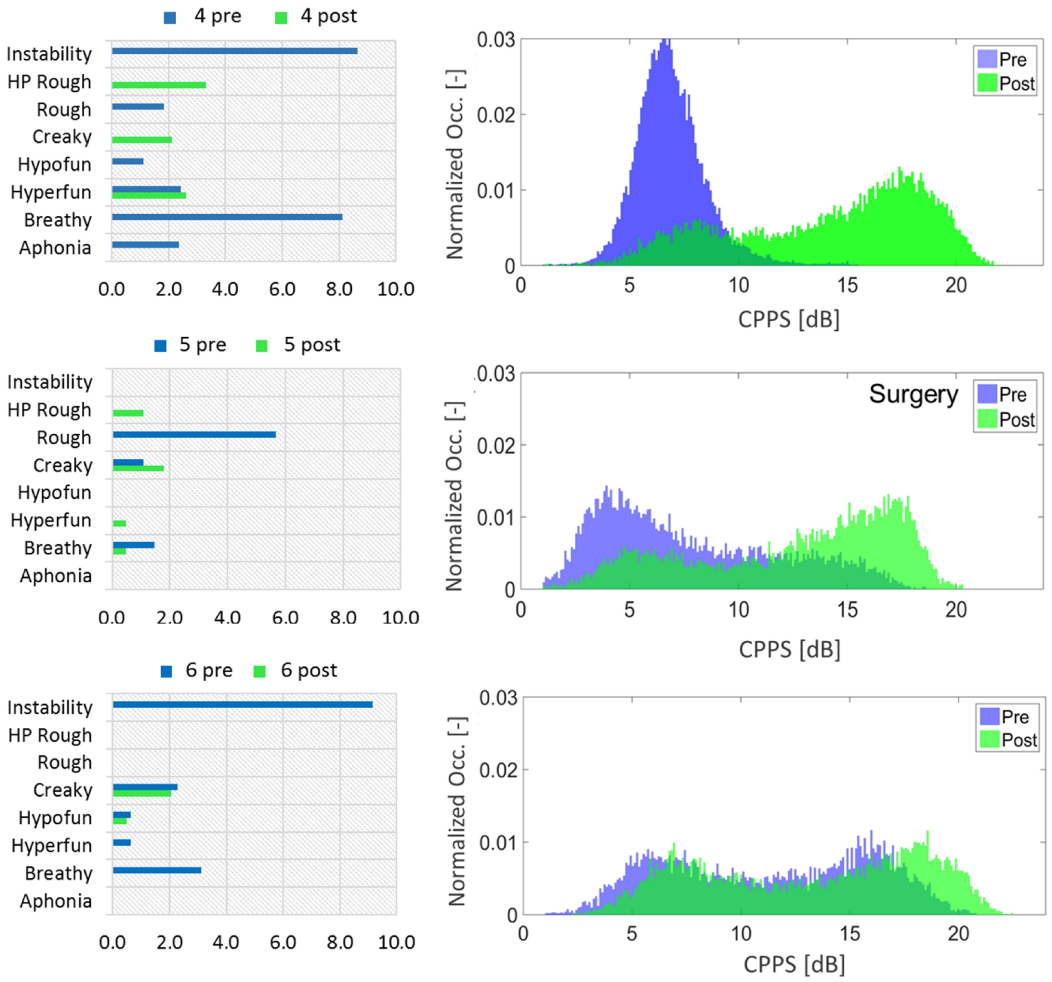

Fig. 6.20 Auditory perceptual evaluation, on the left, and CPPS distributions, on the right, before (blue) and after (green) the speech therapy or the surgery.

Table 6.9 Pearson correlation coefficient between descriptive statistics for CPPS distribution and perceptual ratings (* p-value<0.05; ** p-value<0.01; *** p-value<0.001); *no sig.* not significant p-value.

CPPS parameters	Aphonia	Breathiness	Hyperfunction	Roughness	Instability
$CPPS_{\text{mean}}$	-0.61***	-0.46***	*no sig.*	-0.49***	-0.29*
$CPPS_{\text{median}}$	-0.55***	-0.43***	*no sig.*	-0.54***	-0.34*
$CPPS_{\text{mode}}$	-0.28**	-0.36**	*no sig.*	-0.54***	-0.31*
$CPPS_{\text{std}}$	-0.71***	-0.44***	*no sig.*	-0.28*	*no sig.*
$CPPS_{\text{range}}$	-0.72***	-0.58***	*no sig.*	-0.34*	*no sig.*
$CPPS_{5\text{prc}}$	*no sig.*	-0.25*	*no sig.*	-0.30*	*no sig.*
$CPPS_{95\text{prc}}$	-0.72***	-0.49***	*no sig.*	-0.33**	*no sig.*
$CPPS_{\text{skew}}$	0.43**	0.33*	*no sig.*	0.58***	0.47***
$CPPS_{\text{kurt}}$	0.56***	0.34*	*no sig.*	*no sig.*	0.30*

in the analysis, i.e. "Aphonia", "Breathiness", "Hyperfunction", "Roughness" and "Instability".

Table 6.9 shows that the descriptive statistics for CPPS distribution best correlate with aphonia: the highest correlation coefficient of -0.72 has been obtained for $CPPS_{\text{range}}$ and $CPPS_{95\text{prc}}$, whose negativity indicates that such descriptive statistics decrease as aphonia increases. A strong correlation has been also found between CPPS distribution and both breathiness and roughness: in the case of breathiness $CPPS_{\text{range}}$ and $CPPS_{95\text{prc}}$ have again the highest coefficients of -0.58 and -0.49, while in the case of roughness $CPPS_{\text{skew}}$ and $CPPS_{\text{median}}$ or $CPPS_{\text{mode}}$ show coefficients of 0.58 and -0.54, respectively. Differently from the descriptive statistics that indicates the central tendency of the distribution, $CPPS_{\text{skew}}$ increases when roughness rises, as indicated by the positive Pearson coefficient. The two results are in agreement, since an higher skewness is present when CPPS distribution is centred at very low CPPS values, as for subjects 4 and 5 in Figure 6.20. Few descriptive statistics are correlated with instability (the highest correlation coefficient is equal to 0.47 for $CPPS_{\text{skew}}$), while none of them significantly correlate with hyperfunctional voice. Such outcome was expected, since hyperfunctional voice and strained voice are synonymous and, as already discussed in 6.4 for strained voice, it has differences in spectral harmonics amplitude and not in spectral harmonics periodicity with respect to normal healthy voice.

This work confirms and extends results from previous studies, which investigated the correlation between perceptual ratings and CPPS values obtained from the Hillendbrand software. Brinca *et al.* [87] used the GRBAS scale for the perceptual

assesment of voice and they found the best correlation coefficients between CPPS in readings and breathiness (-0.43), while no significant correlation resulted for the other voice quality features included in the scale (roughness, asthenia and strain). Jannetts *et al.* [189] also used the GRBAS scale and in the case of the reading as speech material, they obtained the highest correlation coefficient of -0.47 for asthenia and comparable coefficients for roughness and breathiness, equalt to -0.35 and -0.38 respectively. Heman-Ackah *et al.* [190] limited the investigation to a sentence and to breathiness and roughness only: they found that both the perceptual aspects significantly correlated with CPPS, with a higher coefficient for breathiness (-0.71). The main limitation in the methodology of the present study is the use of voice samples from Italian and Swedish languages. Despite the need of future works to consider the two languages separately, merging such different phonetic sounds has provided a great variety in the perceptual ratings.

6.4 Study 3: Variability of CPPS distribution in readings of healthy voices

The aim of this study is to provide preliminary normative data to assess results on CPPS distribution for a single subject or a group of speakers. As such, this work recalls the investigations on SPL variability reported in Chapter 3: the same protocol described in paragraphs 4.1.1 and 4.1.2 was followed, but the laboratory and the participants were different.

6.4.1 Method

The experiment was performed in the anechoic chamber of Politecnico di Torino, where the measured A-weighted equivalent background noise level was 26.2 dB. Twenty-five young subjects participated to the experiment, 9 males and 16 females (mean age 23 years, SD of 3.7 years). They were asked to read aloud 2 passages, P1 and P2, twice and in sequence, thus obtaining four repetitions for each subject (for further details see paragraphs 3.1.1 and 3.1.2). The order of readings was randomly changed between subjects, as an improvement in the methodology. The readings were recorded simultaneously by means of three measurement chains, namely:

Fig. 6.21 A subject while performing the reading task.

- the calibrated sound level meter, SLM, with a class 1 omnidirectional measurement microphone. For the entire period of the test, each subject was asked to stand in front of the microphone, on axis, at the fixed distance of 16 cm as provided by a thin spacer (further details in paragraph 3.1.3);
- the omnidirectional headworn microphone, which was placed at a distance of about 2.5 cm from the lips' edges of the talkers, slightly to the side of the mouth, at about 20÷45 degrees horizontally, depending on the subjects' face shape (further details in paragraph 3.1.3);
- the electret condenser microphone, ECM, which was fixed at the jugular notch by means of a surgical band (further details in paragraph 6.3.1);
- the piezoelectric contact microphone, PM, which is embedded in a collar placed around the neck and connected through an AUX cable to the Samsung SM-G310HN smartphone (further details in paragraph 6.3.1).

Figure 6.21 shows a participant while performing the experiment.

Due to incorrect execution of the experiment or temporary unavailability of some devices, a different number of subjects was taken into account for the four devices: 25 subjects (9 males, 16 females) were considered for the SLM and the ECM, 22

subjects (9 males, 13 females) for the headworn microphone, and 20 subjects (7 males, 13 females) for the PM. The wav files collected using the SLM, the headworn microphone and the ECM were down-sampled to 22.05 kSa/s, while the wav files collected by means of the PM already had such sampling rate. A CPPS distribution has been computed for each reading as described in paragraph 6.1, thus obtaining 4 CPPS distributions for each subject. The following descriptive statistics for CPPS distributions have been investigated: mean, $CPPS_{\text{mean}}$, median, $CPPS_{\text{median}}$, mode, $CPPS_{\text{mode}}$, 5th percentile, $CPPS_{5\text{prc}}$, 95th percentile, $CPPS_{95\text{prc}}$, standard deviation, $CPPS_{\text{std}}$, the interval between the maximum and the minimum value, $CPPS_{\text{range}}$, kurtosis, $CPPS_{\text{kurt}}$, and skewness, $CPPS_{\text{skew}}$.

The analyses described in Chapter 3 for SPL measures (paragraphs 3.3.1 and 3.3.2) have been followed for estimating the intra- and inter-speaker variability of each descriptive statistic of CPPS distribution, named as CPPS parameters. A summary of the estimates and the respective symbols is here reported.

Intra-speaker variability:

1. s_{i}: the experimental standard deviation of the four repeated measures for each *i-th* subject;
2. $\overline{s}$ (CI): the average of s_{i} values and its 95% Confidence Interval (CI) for the mean based on a t critical value. It was calculated as 2.09 based on the the sizes of the device-groups [142].

Inter-speaker variability:

1. *s(g)*: the experimental standard deviation of each device-group;
2. s_{m}: the standard deviation of the mean, or standard error;

6.4.2 Analyses and results

Tables 6.10, 6.11, 6.12 and 6.13 show the intra- and inter-speaker variability of CPPS parameters obtained from the readings acquired with each device. They also include the group mean of each CPPS parameter as preliminary normative data for young healthy speakers. As expected, the results obtained for the two microphones in air are

comparable (Tables 6.10 and 6.11). All the CPPS parameters show a limited intra-speaker variability , i.e. $\overline{s}$ lower than 0.3 dB, except $CPPS_{\text{range}}$ and $CPPS_{\text{mode}}$, which have $\overline{s}$ equal to 0.5 dB and 0.8 dB, respectively. The inter-speaker variability results lower than 1 dB for all the cepstral measures, with the exception for $CPPS_{\text{mode}}$, which has *s(g)* of about 2 dB. $CPPS_{95\text{prc}}$, which is the best parameter in discriminating between healthy and pathological voice according to Study 1 (paragraph 6.2), has $\overline{s}$ of 0.18 dB and *s(g)* of 0.42 dB for both the SLM and the headworn microphone. Also the overall group means of the CPPS metrics are comparable for the two microphones in air.

Table 6.10 Results on CPPS variability obtained from the readings recorded with the SLM.

	Intra-speaker		**Inter-speaker**	
CPPS parameter (dB)	$\overline{s}$ (CI)	*s(g)*	s_{m}	group mean
$CPPS_{\text{mean}}$	0.19 (0.15-0.22)	0.70	0.14	13.3
$CPPS_{\text{median}}$	0.24 (0.19-0.29)	0.99	0.20	14.2
$CPPS_{\text{mode}}$	0.76 (0.36-1.15)	2.28	0.46	15.9
$CPPS_{\text{std}}$	0.08 (0.06-0.10)	0.27	0.05	4.4
$CPPS_{\text{range}}$	0.54 (0.44-0.64)	0.80	0.16	20.1
$CPPS_{5\text{prc}}$	0.10 (0.08-0.15)	0.30	0.06	5.4
$CPPS_{95\text{prc}}$	0.18 (0.15-0.22)	0.42	0.08	19.3
$CPPS_{\text{skew}}$	0.06 (0.04-0.07)	0.25	0.31	-0.5
$CPPS_{\text{kurt}}$	0.08 (0.05-0.10)	0.31	0.06	-0.8

Table 6.11 Results on CPPS variability obtained from the readings recorded with the headworn microphone.

	Intra-speaker		**Inter-speaker**	
CPPS parameter (dB)	$\overline{s}$ (CI)	*s(g)*	s_{m}	group mean
$CPPS_{\text{mean}}$	0.20 (0.16-0.24)	0.62	0.13	13.3
$CPPS_{\text{median}}$	0.25 (0.19-0.30)	0.84	0.18	14.2
$CPPS_{\text{mode}}$	0.82 (0.35-1.28)	1.85	0.40	15.7
$CPPS_{\text{std}}$	0.08 (0.06-0.10)	0.21	0.04	4.3
$CPPS_{\text{range}}$	0.50 (0.38-0.61)	0.62	0.13	20.3
$CPPS_{5\text{prc}}$	0.12 (0.09-0.14)	0.29	0.06	5.4
$CPPS_{95\text{prc}}$	0.18 (0.14-0.22)	0.42	0.09	19.3
$CPPS_{\text{skew}}$	0.06 (0.04-0.08)	0.20	0.04	-0.5
$CPPS_{\text{kurt}}$	0.09 (0.06-0.13)	0.27	0.06	-0.7

Differently from the two microphones in air, the two contact microphones show distinct variabilities of CPPS parameters. All the CPPS parameters obtained from the ECM have an intra-speaker variability within 0.4 dB and an inter-speaker variability lower than 1.4 dB (Table 6.12). The intra- and inter-speaker variabilities of CPPS parameters obtained with the PM show instead a similar behaviour of the two microphones in air, but with higher values: Table 6.13 shows that $\overline{s}$ is lower than 0.5 dB for all the metrics with the exception of $CPPS_{\text{range}}$ and $CPPS_{\text{mode}}$, which have $\overline{s}$ equal to 0.7 dB and 0.6 dB, respectively; *s(g)* is higher than 1 dB for 4 out 9 CPPS parameters, with the maximum value of 2.4 dB for $CPPS_{\text{mode}}$.

Table 6.12 Results on CPPS variability obtained from the readings recorded with the ECM.

	Intra-speaker		**Inter-speaker**	
CPPS parameter (dB)	$\overline{s}$ (CI)	*s(g)*	s_{m}	group mean
$CPPS_{\text{mean}}$	0.21 (0.16-0.26)	1.15	0.23	15.5
$CPPS_{\text{median}}$	0.26 (0.20-0.31)	1.37	0.27	16.9
$CPPS_{\text{mode}}$	0.34 (0.27-0.40)	0.82	0.16	19.2
$CPPS_{\text{std}}$	0.13 (0.11-0.15)	0.41	0.08	4.7
$CPPS_{\text{range}}$	0.37 (0.30-0.45)	0.90	0.18	20.7
$CPPS_{5\text{prc}}$	0.23 (0.19-0.28)	0.84	0.17	5.8
$CPPS_{95\text{prc}}$	0.14 (0.11-0.17)	0.69	0.14	20.8
$CPPS_{\text{skew}}$	0.08 (0.06-0.10)	0.32	0.06	-0.9
$CPPS_{\text{kurt}}$	0.19 (0.13-0.25)	0.68	0.14	-0.1

Table 6.13 Results on CPPS variability obtained from the readings recorded with the PM.

	Intra-speaker		Inter-speaker	
CPPS parameter (dB)	$\overline{s}$ (CI)	*s(g)*	s_{m}	group mean
$CPPS_{\text{mean}}$	0.30 (0.16-0.43)	1.13	0.25	12.7
$CPPS_{\text{median}}$	0.41 (0.20-0.61)	1.52	0.34	13.7
$CPPS_{\text{mode}}$	0.71 (0.20-1.22)	2.41	0.54	15.3
$CPPS_{\text{std}}$	0.13 (0.09-0.16)	0.29	0.07	3.9
$CPPS_{\text{range}}$	0.55 (0.45-0.65)	1.10	0.25	17.4
$CPPS_{5\text{prc}}$	0.13 (0.11-0.16)	0.31	0.07	5.3
$CPPS_{95\text{prc}}$	0.25 (0.14-0.35)	0.89	0.20	17.7
$CPPS_{\text{skew}}$	0.10 (0.07-0.14)	0.37	0.08	-0.6
$CPPS_{\text{kurt}}$	0.16 (0.10-0.21)	0.51	0.11	-0.6

Figures 6.22, 6.23, 6.24 and 6.25 show the CPPS distributions obtained from each repetition (a and b) of each passage (P1 and P2) that were performed by two

females and two males. CPPS distributions from the same passage are overlapped, thus highlighting the individual "CPPS vocalprint" again: in the repetitions of two different speech contents the subjects keep their own CPPS distribution. Such an evidence is the basis of the limited intra-speaker variability of all the CPPS parameter, except for $CPPS_{\text{mode}}$ that has higher variability as expected, since it expresses the most frequent value in CPPS distribution.

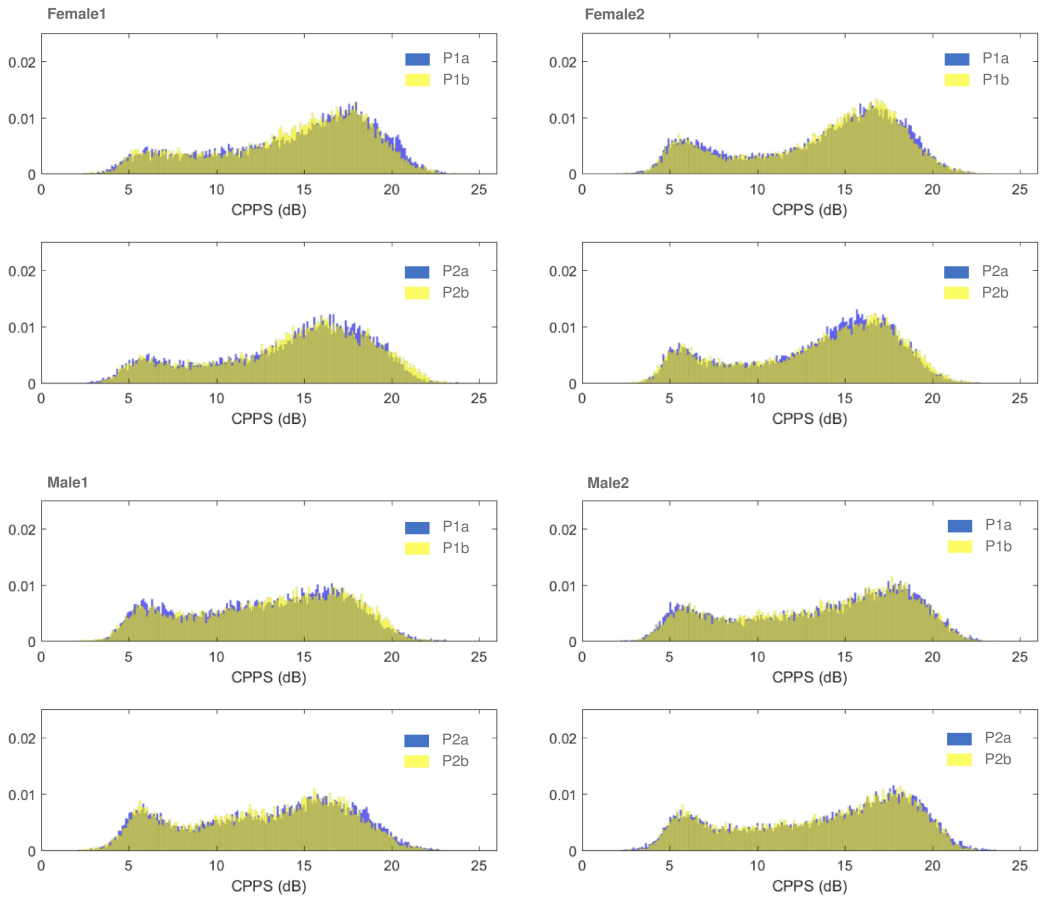

Fig. 6.22 CPPS distributions related to the two readings of the first passage (P1a and P1b) and the second passage (P2a and P2b), acquired with the sound level meter. The distributions belong to two females (upper side) and two males (down side).

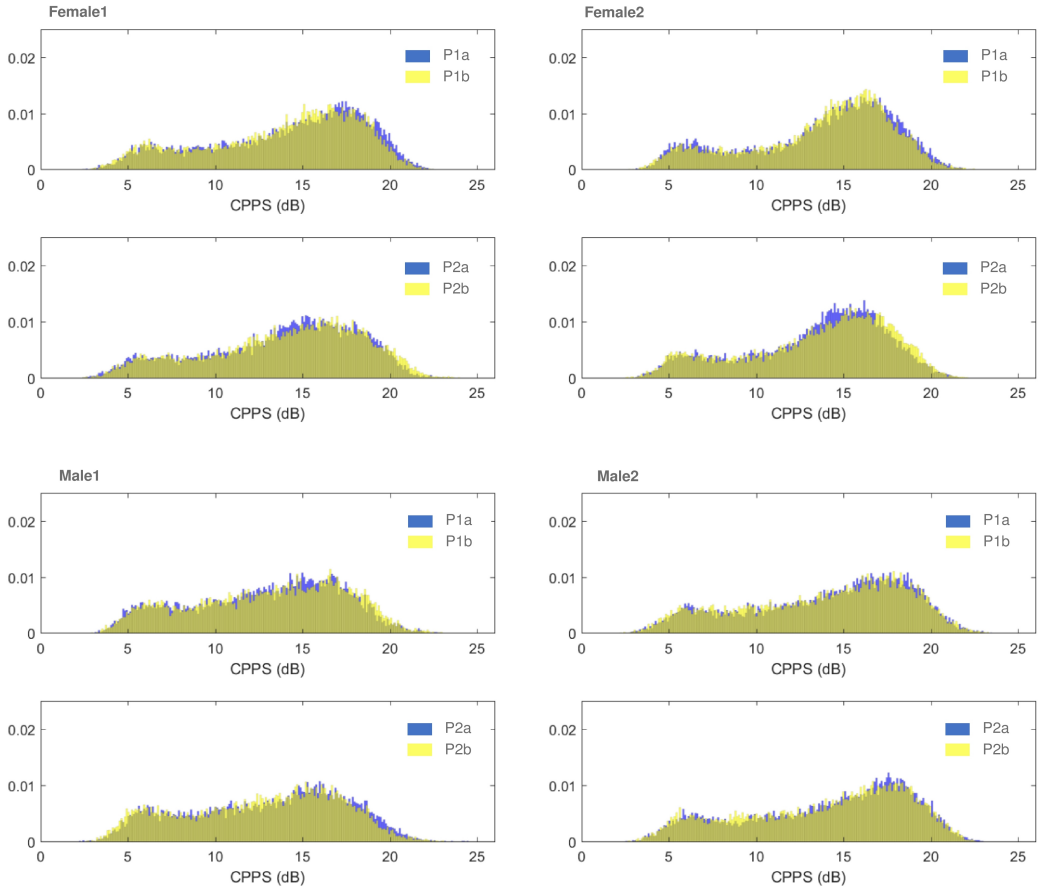

Fig. 6.23 CPPS distributions related to the two readings of the first passage (P1a and P1b) and the second passage (P2a and P2b), acquired with the headworn microphone. The distributions belong to two females (upper side) and two males (down side).

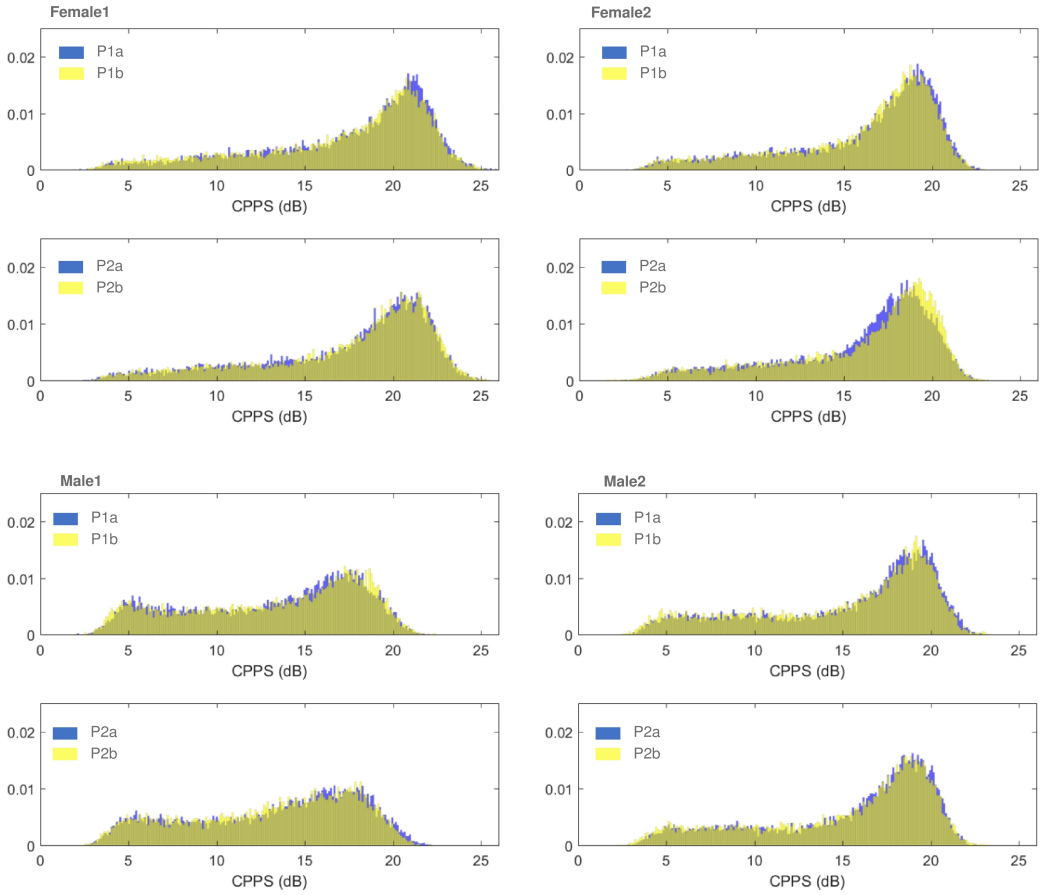

Fig. 6.24 CPPS distributions related to the two readings of the first passage (P1a and P1b) and the second passage (P2a and P2b), acquired with the electret condenser microphone. The distributions belong to two females (upper side) and two males (down side).

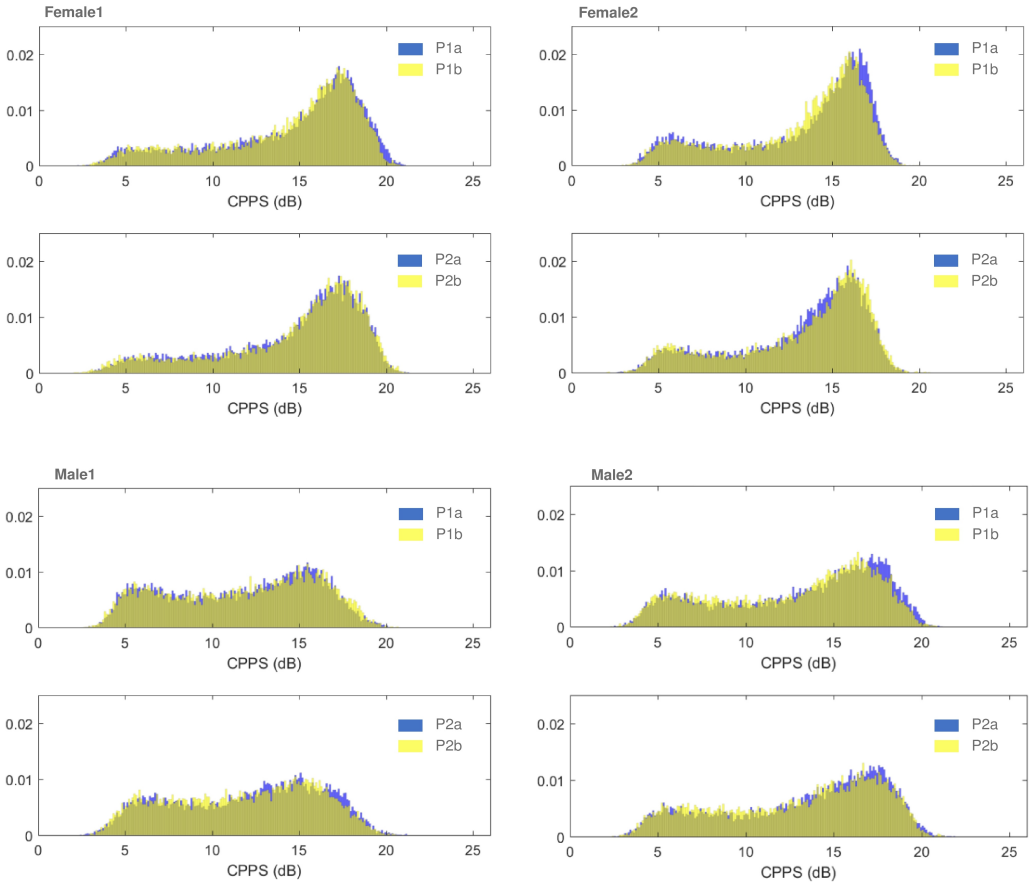

Fig. 6.25 CPPS distributions related to the two readings of the first passage (P1a and P1b) and the second passage (P2a and P2b), acquired with the piezoelectric microphone. The distributions belong to two females (upper side) and two males (down side).

Chapter 7

CPPS and Sample Entropy in vowels excerpted from readings of pathological and healthy speakers

This chapter partially reports material from:

1. A. Castellana, A. Selamtzis, G. Salvi, A. Carullo, A. Astolfi, *Cepstral and entropy analyses in vowels excerpted from continuous speech of dysphonic and control speakers*, in INTERSPEECH 2017, pp. 1814-1818.

In this chapter a further investigation on CPPS distribution is presented, where the speech materials under analysis are the excerpted vowels /a/ from a reading.

Routine clinical examination of vocal health is largely based on perceptual evaluation of the voice quality and videoendoscopic images of the larynx [191]. Although sustained vowels are used for videoendoscopic examination, it has been argued that asking subjects to produce sustained vowels is somehow artificial [191, 183]. For that reason clinicians tend to prefer running speech when they evaluate voice quality perceptually. Based on these evidences, there is the need to focus on a specific widely researched pattern (vowel) taken from its natural context (running speech). Due to the recent spread of non-linear techniques together with cepstral analyses as the most promising measures of voice quality, an additional marker is examined and compared with CPPS in the ability to discriminate between healthy

and pathological voices: the Sample Entropy (SampEn). This chapter aims to address the following questions:

1. How powerful are the CPPS and SampEn metrics in discriminating vocal health in shorter vowels where smoothing in the time dimension is not possible?
2. How do these two metrics correlate with each other?
3. Can the combination of these two metrics improve discrimination of vocal health?

7.1 Voice samples

The study sample consisted of 33 voluntary patients, 25 females and 8 males (age range: 20-82 years; mean: 50.0 years; standard deviation SD: 16.5 years) and 31 healthy adults with normal voices, 18 females and 13 males (age range: 19-49 years; mean: 25.4 years; SD: 7.8 years). All subjects were native Italian speakers.

Each participant followed the same procedure, which can be summarized in two steps:

1. They read aloud an Italian standardized phonetically balanced passage of 300 words at a comfortable pitch and loudness. It tooks an average reading time of about 2 minutes. The reading text is the first reported in the appendix;
2. Two otolaryngologists performed the clinical practice that included a careful case history and the videolaryngoscopy examination.

Table 7.1 summarizes the otolaryngologic diagnoses and their occurrences in the patient group. Before performing step (1) of the protocol, subjects worn an omni-directional head mounted microphone Mipro MU-55HN, which was placed at a distance of about 2.5 cm from the lips of the speaker, slightly to the side of the mouth. Details about the microphone have been described in Chapter 3, paragraph 3.1.3. Voice recordings were performed in a quiet room with an A-weighted equivalent background noise level of 50.0 dB (SD = 2 dB).

Table 7.1 Diagnoses for the patient group.

Organic dysphonia	Number
Cyst	7
Edema	9
Sulcus vocalis	3
Polyp	4
Chronic laryngitis	2
Vocal fold hypostenia	3
Vocal fold paresis	2
Vocal fold nodul	1
Post-surgery dysphonia	2

7.2 Data processing

The first 118 words of each reading were considered in our analyses. Phonemic annotations for the speech files were created using automatic speech recognition and the orthographic transcriptions (prompts) by means of forced alignment. The recognizer was trained on the Italian SpeechDat corpus using HTK [192] and the RefRec scripts [193]. The SpeechDat corpus includes more than 3000 speakers recorded over the telephone line with 8-bit, 8 kHz, A-law quality. Rather than resampling our speech files to fit the acoustic models, the feature extraction procedure was modified to limit the mel scale filterbank from 0 to 4 kHz. Monophone models with 32 Gaussian components per HMM state were used. Out-of-vocabulary words and their canonical pronunciation were added to the dictionary. This step does not affect the automatic nature of the method because the prompt text in this application is known in advance. The time-aligned transcriptions were used to extract speech samples belonging to all occurrences of the /a/ vowel both for pathological and control speakers.

For computational reasons, signals were down-sampled to 25 kHz the middle 1024 samples (40.96 ms) were extracted from each vowel waveform to consider a segment unaffected by transient onset and offset behavior. For each vowel, a CPPS measure and a SampEn measure were computed, thus obtaining a time series of each parameter per subject, which was treated as a distribution. The size of the time series ranges from 43 to 71 values. For each individual distribution the following

descriptive statistics were calculated: mean, median, standard deviation (std), range, 5th percentile ($5prc$) and 95th percentile ($95prc$).

7.3 Metrics

7.3.1 CPPS

An adjusted version of the Cepstral Peak Prominence Smoothed (CPPS) algorithm described in 5.1 has been implemented: since vowel segments excerpted from continuous speech are not long enough, CPPS computation has not included the time-smoothing step of cepstra.

7.3.2 SampEn

Sample Entropy (SampEn) is a metric from the field of nonlinear time series analysis, introduced by [194], as the successor of Approximate Entropy (ApEn) [195]. SampEn quantifies the irregularity of a time series; a low number for SampEn signifies regularity while a higher number signifies increasing degree of irregularity. This means that for a healthy voice signal the expected SampEn value is low while for a pathological voice the presence of irregularity should result in a higher SampEn value. A number of studies have used SampEn and ApEn for assessing voice irregularity from electroglottographic or acoustic signals [196, 197].

SampEn is calculated [194, 198] by separating a time series into sequences of length of m and m+1 points. Then the conditional probability is calculated that the Chebyshev distance between two sequences of length m+1 is less than a tolerance r, given that the Chebyshev distance between two sequences of length m is less than a tolerance r. This probability is calculated as the ratio of the number of sequences of length m+1 whose Chebyshev distances are less than r over the number of sequences of length m whose Chebyshev distances are less than r. SampEn is then defined as the negative natural logarithm of this ratio.

For each vowel, SampEn was calculated for a number of overlapping windows of length N, using the MATLAB implementation provided in [198]. The length of each window N was f0 dependent and equal to four times the length of the glottal cycle,

while the overlap was equal to the length of one glottal cycle. The f0 was estimated using the YIN algorithm [199]. The reason for adapting the window length in this way is the sensitivity of SampEn to fundamental frequency [200] and maintaining consistency of the analyzed pattern across subjects. The reader should keep in mind that this type of windowing is best applicable for Type 1 and Type 2 (nonchaotic) voice signals [201]. The total estimate of SampEn per vowel was taken as the mean of the values obtained from all windows. As suggested in [197], the sequence length m was taken to be equal to round(log10(N)) and the matching tolerance r was taken to be 0.1 times the standard deviation of the N points in the analyzed window.

7.4 Statistical analyses

The two-tailed Mann-Whitney U-test, a non-parametric test based on independent samples [149], was used to evaluate statistical differences of the paired lists of descriptive statistics related to the groups of healthy and pathological subjects. The null hypothesis states that MD = 0, where MD is the median of the population of the differences between the sample data for the two groups. If the null hypothesis is accepted, the two lists of values come from the same population, i.e. it is not possible to distinguish healthy and pathological voice samples. The selection of the Mann-Whitney U-test was made based on the one-sample Kolmogorov-Smirnov test, which verified that the data in each list did not come from a normal distribution. The Spearman correlation coefficient was used to determine the relationship between CPPS and SampEn values. The above-mentioned tests were performed using a MATLAB script (R2014b, version 8.4).

The Receiver Operating Characteristic (ROC) analysis was used to investigate the discrimination power of each descriptive statistic of CPPS and SampEn distributions in healthy and pathological voices. The area under the Receiver Operating Characteristic, which is named Area Under Curve (AUC), was calculated as an indicator of classification accuracy [202]. The AUC ranges from 0.5 to 1.0: a value higher than 0.9 indicates an outstanding separation between the two groups, an AUC between 0.9 and 0.8 designates an excellent discrimination power, an AUC between 0.8 and 0.7 an acceptable discrimination, while an AUC close to 0.5 shows a poor ability to separate the two groups. Moreover, the ROC analysis was performed to determine the preliminary criteria for positivity, i.e., the threshold value from which

a voice could be indicated as pathological, for CPPS and SampEn in [a] vowels excerpted from continuous speech. The optimal threshold between the dysphonic group and controls was evaluated plotting together sensitivity and specificity versus each possible threshold. Sensitivity corresponds to the true positive rate, i.e. the percentage of subjects with voice disorders that are classified as positive. Specificity is the true negative rate, that is the percentage of people with healthy voice who are identified as negative. Instead of conventionally selecting the optimal threshold value where the two curves cross, the authors gave priority to the sensitivity, maximizing in this way the proportion of subjects with voice disorders that are classified as positive. As a final step, a logistic regression model combining both cepstral and entropy measures as predictive variables and assuming the presence/absence of dysphonia as dependent variable was performed. The Wald test was carried out for assessing the significance of coefficients for each predictive variable, where the null hypothesis states that the coefficient of the variable is equal to zero, i.e., the variable is not contributing to the logistic model [175]. These analyses were performed using the statistical program SPSS (v. 21; SPSS Inc, New 223 York, NY).

7.5 Results

Table 7.2 shows the p-values of the Two-tailed Mann-Whitney U-test of the lists of descriptive statistics for CPPS distributions related to the patients and controls. The p-values were lower than 0.05 for the *mean*, *median*, $5prc$ and $95prc$, which means the null hypothesis was rejected. The *std* and the *range*, instead, had p-values higher than 0.05. These outcomes reveal that CPPS distributions are significantly different in central tendency, with an overall average value of 12.2 dB and 14.1 dB for the mean in patients and controls, respectively, but not in variance, with an overall average value of 3.0 dB and 2.8 dB for the std in patients and controls, respectively. Table 7.2 also shows that only the *mean* and $95prc$ of the SampEn distributions had p-values of the two-tailed Mann-Whitney U-test lower than 0.05. SampEn distributions are thus significantly different in central tendency, with an overall average value of 0.7 and 0.5 for the *mean* in patients and controls, respectively, but not in variance, having an overall average value of 0.3 and 0.2 for the *std* in patients and controls, respectively.

Table 7.2 Analysis results for each paired list of descriptive statistics related to the dysphonic group and the control group. Two-tailed Mann-Whitney U-test *p*-values: values lower than 0.05 are in bold and indicate the rejection of the null hypothesis. Area Under Curve (AUC) and the relative 95% Confidence Interval (CI).

Statistics	**CPPS**		**SampEn**	
	U-test	AUC (CI)	U-test	AUC (CI)
Mean	**0.001**	0.85 (0.76;0.95)	**0.002**	0.72 (0.59;0.85)
Median	**0.001**	0.85 (0.75;0.94)	0.999	0.70 (0.58;0.84)
Std	0.116	0.59 (0.45;0.73)	0.884	0.59 (0.45;0.73)
Range	0.158	0.61 (0.47;0.74)	0.509	0.45 (0.31;0.60)
5prc	**0.001**	0.73 (0.61;0.85)	0.993	0.67 (0.54;0.80)
95prc	**0.001**	0.79 (0.68;0.90)	**0.004**	0.71 (0.58;0.84)

In Table 7.2 the discrimination power of each descriptive statistic for CPPS and SampEn distributions is highlighted through AUC values and the respective 95% Confidence Intervals (CI). Generally, all the descriptive statistics from CPPS distributions had higher AUCs than the ones from SampEn distributions, highlighting a better diagnostic precision of the cepstral measure with respect to the entropy metric. Among the descriptive statistics from CPPS distributions, the mean and the median had an AUC of 0.85, showing a good discrimination power between dysphonic and healthy subjects. Regarding the descriptive statistics from SampEn distributions instead, the mean had the highest AUC of 0.72, highlighting a moderate discrimination power. Figure 7.1 shows the boxplots of the *mean* from individual CPPS distributions for the group of patients and controls: as expected, healthy speakers had higher values than patients. The bold line in Figure 1 indicates the optimal threshold obtained from the ROC analysis for the mean, i.e., 14.0 dB. The criterion for positivity thus corresponds to 14.0 dB or lower, with a sensitivity of 79% and a specificity of 71%.

Figure 7.2 shows the boxplots of the *mean* from SampEn distributions for subjects belonging to the healthy group and the pathologic one: predictably, an opposite trend with respect to Figure 1 is evident, having higher values of entropy for subjects with voice disorders. The criterion for positivity is 0.58 or higher, with a sensitivity of 73% and a specificity of 68%.

A Spearman correlation coefficient of -0.61 (p-value<0.001) was found between CPPS *mean* and SampEn *mean*, thus highlighting a strong negative correlation

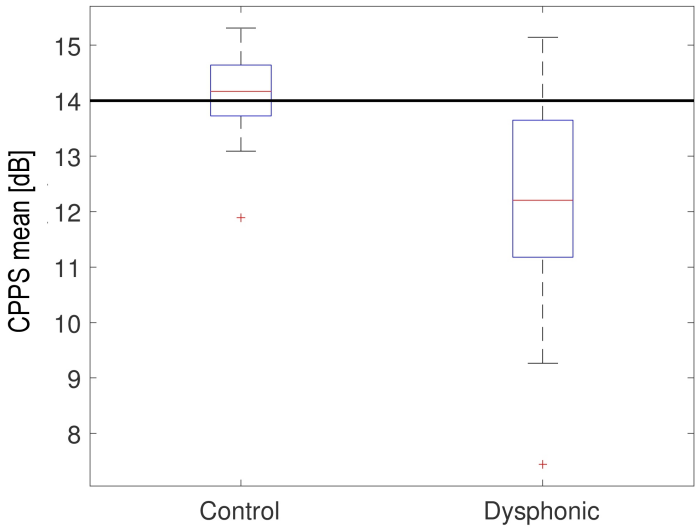

Fig. 7.1 Boxplots of the mean from individual CPPS distributions for the controls and the dysphonic group. The bold line indicates the best threshold of 14.0 dB.

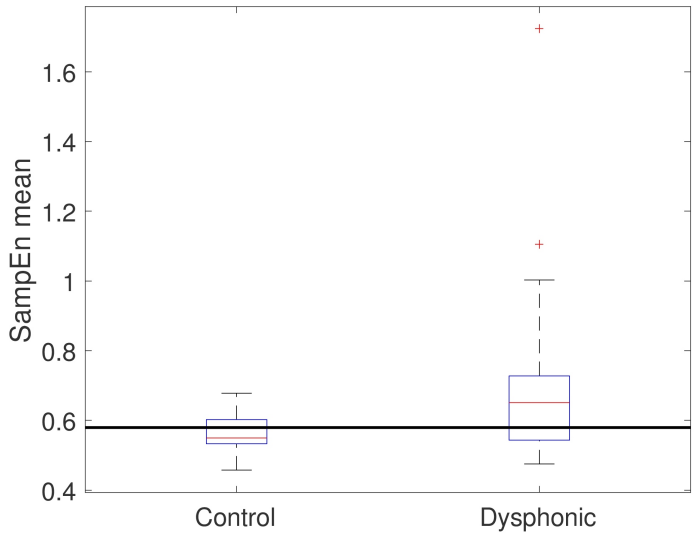

Fig. 7.2 Boxplots of the *mean* from individual SampEn distributions for the controls and the dysphonic group. The bold line indicates the best threshold of 0.58.

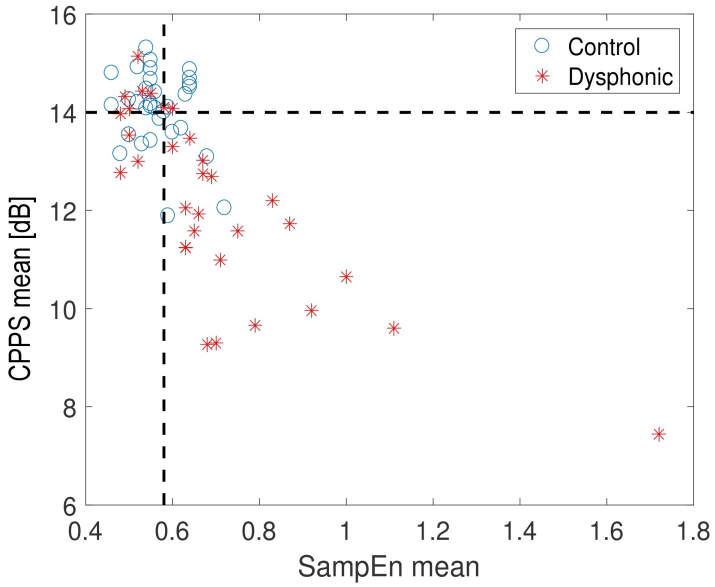

Fig. 7.3 Scatter plot between CPPS *mean* and SampEn *mean*: circles indicate the controls; stars represent the dysphonic group. The dashed lines indicate the best thresholds for the two parameters (14 dB for CPPS and 0.58 for SampEn).

between the two parameters, i.e. when one increases the other decreases (see Figure 7.3). The Wald test p-values in the output of the logistic regression model with CPPS *mean* and SampEn *mean* as predictive variables were 0.002 and 0.874, respectively, underlying that only CPPS mean is a significant variable in predicting dysphonia for the present database.

The limited number of subjects with a given pathology did not allow for a statistical testing that would reveal which pathologies are easier to discriminate using CPPS or SampEn. However, by checking the CPPS and SampEn results in relation to the diagnosis, it was observed that some pathologies were more successfully discriminated than others. CPPS mean and SampEn mean discriminated successfully vocal fold paresis (2 of 2), polyp (4 of 4), and post-surgery dysphonia (2 of 2). Moreover, CPPS mean discriminated laryngitis (2 of 2) and it partially differentiates edema (5 of 9), cyst (3 of 5) and sulcus (2 of 3). Future studies need to consider homogeneous data (in gender and diagnosis) and larger data samples that include the perceptual assessment of voice, in order to investigate correlations between the parameters and the degree of dysphonia. Additionally, the investigation of the metrics in several languages and the study of the relationships with the perceptual

evaluations of voice would improve the applicability of the analysis in different contexts. The automatic method and the preliminary results of this study support in-clinic voice monitoring for the characterization of vocal health and self-monitoring of voice during everyday activities to characterize changes in the vocal behavior of professional voice users.

Chapter 8

Conclusions and future directions

In this final chapter, a summary of the main findings based on the objectives of this Ph.D. research is presented and considerations on the possible future applications of the obtained outcomes are given.

Objective 1: **To investigate teachers' vocal behaviour and to study the relationships between voice use and classroom acoustic parameters, through in-field longitudinal observations over a school year.**

Chapter 2 presents the investigations on changes in the voice use of secondary school teachers and in the activity background noise conditions over a period of one school year, as well as on the relationship between the teachers' voice parameters and the acoustic conditions inside classrooms. The vocal activity of 31 and 22 teachers from two secondary schools was monitored for two working weeks at the beginning (stage 1) and at the end (stage 2) of the same school year using a contact-sensor based vocal analyzer, namely Voice Care. The teachers' voice parameters acquired during the two stages were compared and analyzed in relation to the measured classroom acoustic parameters of reverberation time $T30_{0.25\div2\text{kHz,occ}}$ and background noise level L_{A90}.

Teachers spoke with a higher sound pressure level of voice ($SPL_{\text{mean,1m}}$) at the end of a school year during conversational tasks (average increase equal to 3.8 dB), thus highlighting that the use of a high voice level during working activities makes teachers also use a higher voice level during non-occupational activities.

A significant variation of some vocal parameters was observed during the teaching activities, but only in the school with the highest noise and reverberation time values. The teachers in this school showed an increase in $SPL_{\mathrm{mean,1m}}$ and a reduction in the voicing time percentage (Dt%) at the end of the school year. In summary, teachers who work in poor acoustic conditions use a higher level of voice during working activities at the end of the school year, and they probably decrease Dt% in order to reduce the feeling of fatigue due to the use of high voice levels.

The measurements of the activity background noise level L_{A90} revealed a significant increase in noise at the end of the school year. This indicates that, after one year of exposition to high noise levels, students tend to make more noise at the end of the school year, possibly because of their feelings of fatigue. Future studies could be performed in order to assess whether students perceive such increase. In this way, it would be possible to verify whether, after a long exposition to high noise levels, people are better able to tolerate high levels of noise and consequently make more noise.

Finally, the positive association between L_{A90} and reverberation both at the beginning and at the end of the school year has confirmed that background noise, in occupied conditions, is affected by the reflections that are present in the room, and that an acoustical renovation would be necessary to reduce noise levels.

As far as the influence of the background noise level on occupational voice parameters is concerned, a *Lombard effect* was found at a rate of 0.4 dB/dB at the beginning of the school year. At the end of the school year, it was found that teachers spoke with higher voice levels and that noise did not seem to have a significant influence on the variations of their voice level. It could be interesting to investigate whether an upper limit to the Lombard effect exists, which each teacher has difficulty in overcoming. When studying the relationship between occupational voice parameters and reverberation, it was observed that $SPL_{\mathrm{mean,1m}}$ was related to the $T30_{0.25\div 2\mathrm{kHz,occ}}$ values through a quadratic regression curve, with the minimum value of SPL at about 0.8 s of $T30_{0.25\div 2\mathrm{kHz,occ}}$ in both stages. A reverberation time of 0.8 s is in good agreement with other studies that indicated similar values to guarantee support to voice of the speaker and vocal comfort.

A good-fit model was found that allows the speech sound pressure level, at 1 m in front of the speaker's mouth, to be predicted from the background noise level and the mid frequency reverberation of the room, taking into account the within

subject dependence of the voice parameters. Furthermore, this analysis indicated an inter-subjects variability (standard deviation) of SPL between subjects in the same surrounding conditions of 3.5 dB, and a variability of the SPL around the individual regression line of 3.0 dB. Although the overall amount of vocal data collected in this study was sufficient to provide a good model to estimate the sound pressure level in front of the speaker's mouth, future research should consider a larger database in order to achieve a better validation of the model. In short, since this is the first longitudinal study over one school year and no data exists that would allow the results to be compared, further research could be conducted to verify the main results of the present work.

Objective 2: **To investigate differences in speech intensity in very low and very high reverberant rooms, accounting for the uncertainty of the parameters estimated using a headworn microphone and a vocal analyzer.**

Chapter 3 deals with the variability of speech SPL within a speaker, i.e., intra-speaker variability, and in a group of people, i.e., inter-speaker variability, in successive readings that were recorded with a sound level meter, a headworn microphone, and Voice Care. For each device, the intra-speaker variability was within 1 dB for SPL_{eq} and SPL_{mean}, while it reached 2 dB for SPL_{mode}; the inter-speaker variability ranged from 2.8 to 5.3 dB, having always the highest values for SPL_{mode}. In addition, the intra-speaker variability was always lower than the respective inter-speaker variability. For the sound level meter and the headworn microphone, negligible changes of descriptive statistics for SPL distributions and SPL_{eq} were obtained in the repetition of the same passage, while significant differences were found in readings of different passages. Fewer modifications were highlighted for Voice Care, probably due to the different post-processing that it implements for the SPL estimation. Both the intra- and inter-speaker variability of SPL_{eq} remained constant as logging interval changes, while SPL_{mean} and SPL_{mode} showed moderate to high sensitivity with respect to the logging interval used in the post-processing.

The first part of **Chapter 4** provides the uncertainty that affects instantaneous speech SPL, absolute values and differences between SPL parameters, such as equivalent, SPL_{eq}, mean, SPL_{m}, and mode, SPL_{mode}. Speech measurements were performed with Voice Care and with the headworn microphone Mipro MU-55HN. Assuming that an accuracy of ± 3 dB can be considered acceptable for absolute SPL parameters according to Schutte and Seidner [146], both Voice Care and the

headworn microphone comply with this requirement, except for the SPL_{mode}. It is advisable the usage of an headworn microphone, which showed an uncertainty of ≈ 2 dB for the most used parameter SPL_{eq}, instead of a voice monitoring device equipped with a contact microphone fixed at the jugular notch, for which a correspondent value of ≈ 3 dB was obtained. However, in situations when a microphone in air is not suitable e.g. high background noise levels or long-term voice monitorings, the advantage in using a contact microphone is doubtfulness more evident despite its higher SPL uncertainty, which users should state in the results delivery. On the other side, comparable uncertainties have been shown for the two devices when SPL_{m} differences between two speeches have to be assessed, making them interchangeable in speech monitoring.

The second part of **Chapter 4** deals with the effect of very low and excessive reverberation on SPL of continuous speech. Measurements were carried out in a semi-anechoic and reverberant room using Voice Care and the MU-55HN headworn microphone, placed at 2.5 cm by the speaker's mouth. University students evoked short monologues in which they explained something they knew well and also described a map with the intent of correctly explaining directions to a listener who drew the path on a blank chart 6 m away. The uncertainty estimation of SPL parameters was also taken into account for the results assessment. Concerning the recordings with Voice Care, which uses a 30 ms frame length for selecting voiced frames only, a statistically significant increase of about 2 dB in the overall average of SPL_{eq}, SPL_{m} and SPL_{mode} in the semi-anechoic room compared to the reverberant room was found for the map description, thus highlighting an increasing vocal effort in a dead room with a speech task that requires a communicative intent. The same behaviour was not obtained for the free speech task. In the case of the headworn microphone, for which a logging interval of 1 s was used, no significant differences were found neither in speech sound pressure levels nor in sound power levels. Investigations were carried out for determining how a 30 ms logging interval, whose length is comparable to the inter-syllabic pause, can affect the SPL parameters estimation while using the headworn microphone. A 1 s logging interval is suggested for the estimation of SPL parameters with headworn microphones, as it makes them less affected by the noise recorded in the speech pauses. In particular, the less influenced parameter is SPL_{eq}. However, this study showed as contact-based microphone devices are more appropriate than microphones in air for speech monitoring, because they are able to detect the vocal-fold activity only, without any noise artefacts.

Objective 3: To validate CPPS distributions as vocal health indicator in sustained vowels and continuous speech using microphones in air and contact ones.

Three chapters of this thesis explore the Cepstral Peak Prominence Smoothed (CPPS) distributions as descriptors of vocal health status in different speech materials.

Chapter 5 investigates descriptive statistics for CPPS distribution in sustained vowels /a/ vocalized by 41 patients and 35 controls and simultaneously acquired with a headworn microphone and a contact sensor, namely an Electret Condenser Microphone (ECM). Regarding the vowels acquired with the microphone in air, the fifth percentile ($CPPS_{5\text{prc}}$) resulted the best descriptive statistic for CPPS distributions that is able to discriminate healthy and unhealthy voices. The respective empirical logistic model showed a strong discrimination power (Area Under the Curve, AUC = 0.95) and a discrimination threshold of $CPPS_{5\text{prc}}$=15.0 dB (95% Confidence Interval, CI, of 0.7 dB), with lower values indicating unhealthy status of voice. Concerning the sustained vowels acquired with the ECM, instead, the standard deviation ($CPPS_{\text{std}}$) was the best parameter that separates the two groups. The respective empirical logistic model had a good discrimination power, with AUC of 0.87, and a discrimination threshold of $CPPS_{\text{std}}$=1.1 dB (95% CI of 0.2 dB), with larger values for pathological voice. Differently from the results by Mehta *et al.*[92], the proposed method is able to discriminate healthy and unhealthy voice from both the microphone in air and a contact microphone. The intra-speaker variability of the two CPPS parameters was larger in the patients group than in the control one: its respective values were 0.8 dB and 0.5 dB for $CPPS_{5\text{prc}}$ and 0.3 dB and 0.2 dB for $CPPS_{\text{std}}$. This result highlights the limited capability of patients in the vocal production.

With the aim of providing guidelines that make the estimated CPPS parameters reliable, an analysis of the main CPPS influence quantities was performed. The obtained outcomes highlighted that the fundamental frequency and the Signal-to-Noise Ratio (SNR) level of the acquired signals could significantly affect the discrimination between healthy and pathological voices. For this reason, it is important to limit the field of use of the fundamental frequency, e.g. providing a reference tone to the subject before he/she performs the speech task, and to avoid large difference in the SNR level during the experimental campaign. Further investigations were made in order to estimate the effect of the frequency content of the signal spectrum on

the CPPS parameters. As the result of this analysis, it can be stated that a reliable estimation of the parameters $CPPS_{5\text{prc}}$ and $CPPS_{\text{std}}$ is obtained provided that the frequency content of the spectrum is not lower than 5 kHz. This justifies the lower discrimination power obtained for the contact microphone that showed a frequency content of about 3.5 kHz.

Chapter 6 investigates descriptive statistics for CPPS distribution in continuous speeches performed by 72 patients and 39 controls and simultaneously acquired with a headworn microphone and a contact sensor, namely an Electret Condenser Microphone (ECM). The 95th percentile of CPPS distributions, $CPPS_{95\text{prc}}$, was the best CPPS parameter in discriminating between healthy and pathological voices, for both reading and free speech tasks (AUC of 0.86). The discrimination threshold was equal to 18 dB (95% CI of 0.6 dB), where lower values indicate a high probability to have unhealthy voice. Concerning the speech materials acquired with the ECM, a reasonable discrimination power (AUC higher than 0.80) was not obtained in the case of continuous speech.

A strong within consistency across tasks was found for both the types of sensor: the central tendency and the shape of individual CPPS distributions were kept in reading and free speech tasks. Such characteristic was indicated as individual “CPPS vocalprint”.

Moderate correlations resulted between $CPPS_{95\text{prc}}$ from reading and free speech and the scores of voice self-assessment performed using the *Profilo di Attività e Partecipazione Vocale* (PAPV).

Changes in CPPS distributions in different voice qualities, i.e. "normal", "creaky", "breathy" and "strained" voice, produced by 5 voice experts while reading were also investigated. The speech samples were simultaneously acquired with a microphone in air and two contact sensors in a sound-booth, namely an ECM and a piezoelectric microphone. CPPS distributions appeared to have different shape and central tendency for each voice quality, although it was confirmed that CPPS distributions varied with the characteristics of the measurement chain, i.e. inner noise floor and bandwidth. Between the two contact sensors, the piezoelectric microphone had a higher frequency content than the ECM, which reaches 5 kHz, thus showing a bandwidth for reliable CPPS measures, as described in Chapter 5. However, most of the CPPS parameters were strongly correlated between the three devices.

Furthermore, 2 expert speech pathologists performed the auditory perceptual rating in consensus over 20 Swedish readings and 36 Italian readings acquired with a microphone in air: most of CPPS parameters showed moderate-to strong correlation with aphonia, breathiness and roughness. These findings are corroborated by the study on 6 patients who read a passage before and after voice therapy or phono-surgery: when aphonia, breathiness or roughness disappeared after the intervention, CPPS distributions significantly changed. Therefore, CPPS distribution may help in the diagnostic procedure and furnish proof of outcomes after interventions.

Future researches need to use larger and more homogeneous data samples in order to investigate on CPPS distributions using clustering methods. Moreover, the diagnostic precision of CPPS parameters will also be stated for voice samples acquired with the piezoelectric microphone, that seems a promising contact microphone for this kind of voice quality estimation. Other types of contact microphone will also be tested. In-field long-term monitorings of teacher's voice using a smartphone application combined with a cheap contact microphone embedded in a collar, the Vocal Holter App (PR.O.VOICE, Turin, Italy), are also planned. In this way, CPPS distributions will be investigated in a real context and their visual feedback for occupational voice users will be tested.

Chapter 7 presents a preliminary study that investigates the efficacy of Cepstral Peak Prominence Smoothed (CPPS) and Sample Entropy (SampEn) in discriminating between dysphonic and healthy subjects using excerpted vowels from readings acquired with a headworn microphone. The mean from both CPPS and SampEn distributions resulted the best in discriminating the two groups, but CPPS mean had higher diagnostic precision than SampEn mean (Area Under Curve of 0.85 and 0.72, respectively). A strong negative correlation of -0.61 was also found between the two metrics, with higher SampEn corresponding to lower CPPS. Worse voice quality results, in fact, in lower values for SampEn and higher values for CPPS, since the first represents the grade of disorder of the vocal signal in the time domain, while the second denotes the regularity of harmonics in the spectrum. The strong relationship between the two measures highlighted in this work is now conducting to further investigations on the effect of vowel context: our aim is to observe the diagnostic precision of cepstral and entropy analysis from healthy and pathological voices in both excerpted and sustained vowels.

Future studies need to consider homogeneous data (in gender and diagnosis) and larger data samples that include the perceptual assessment of voice, in order to investigate correlations between the parameters and the degree of dysphonia. The automatic method and the preliminary results of this study support in-clinic voice monitoring for the characterization of vocal health and self-monitoring of voice during everyday activities to characterize changes in the vocal behaviour of professional voice users.

References

[1] M Behlau. Organizer. voz: o livro do especialista. 2001.

[2] Erkki Vilkman. Voice problems at work: a challenge for occupational safety and health arrangement. *Folia phoniatrica et logopaedica*, 52(1-3):120–125, 2000.

[3] Mara Behlau, Fabiana Zambon, Ana Cláudia Guerrieri, and Nelson Roy. Epidemiology of voice disorders in teachers and nonteachers in brazil: prevalence and adverse effects. *Journal of voice*, 26(5):665–e9, 2012.

[4] Pere Godall, Cecília Gassull, Anna Godoy, and Miquel Amador. Epidemiological voice health map of the teaching population of granollers (barcelona) developed from the eves questionnaire and the vhi. *Logopedics Phoniatrics Vocology*, 40(4):171–178, 2015.

[5] Nelson Roy, Ray M Merrill, Susan Thibeault, Rahul A Parsa, Steven D Gray, and Elaine M Smith. Prevalence of voice disorders in teachers and the general population. *Journal of Speech, Language, and Hearing Research*, 47(2):281–293, 2004.

[6] Erkki Vilkman. Occupational safety and health aspects of voice and speech professions. *Folia Phoniatrica et Logopaedica*, 56(4):220–253, 2004.

[7] R Buekers, E Bierens, H Kingma, and EHMA Marres. Vocal load as measured by the voice accumulator. *Folia phoniatrica et logopaedica*, 47(5):252–261, 1995.

[8] Susanna Whitling, Roland Rydell, and Viveka Lyberg Åhlander. Design of a clinical vocal loading test with long-time measurement of voice. *Journal of Voice*, 29(2):261–e13, 2015.

[9] International Organization for Standardization, Geneva, Switzerland. *ISO 9921, Ergonomics—Assesssment of Speech Communication*, 2003.

[10] Ingo R Titze, Jan G Svec, and Peter S Popolo. Vocal dose measuresquantifying accumulated vibration exposure in vocal fold tissues. *Journal of Speech, Language, and Hearing Research*, 46(4):919–932, 2003.

[11] A Nacci, B Fattori, V Mancini, E Panicucci, F Ursino, FM Cartaino, and S Berrettini. The use and role of the ambulatory phonation monitor (apm) in voice assessment. *ACTA otorhinolaryngologica italica*, 33(1):49, 2013.

[12] Mara Behlau, Fabiana Zambon, and Glaucya Madazio. Managing dysphonia in occupational voice users. *Current opinion in otolaryngology & head and neck surgery*, 22(3):188–194, 2014.

[13] Ingo R Titze, Julie Lemke, and Doug Montequin. Populations in the us workforce who rely on voice as a primary tool of trade: a preliminary report. *Journal of Voice*, 11(3):254–259, 1997.

[14] Elaine Smith, Jon Lemke, Margaretta Taylor, H Lester Kirchner, and Henry Hoffman. Frequency of voice problems among teachers and other occupations. *Journal of voice*, 12(4):480–488, 1998.

[15] Adriane Mesquita de Medeiros, Ada Ávila Assunção, and Sandhi Maria Barreto. Absenteeism due to voice disorders in female teachers: a public health problem. *International archives of occupational and environmental health*, 85(8):853–864, 2012.

[16] Lady Catherine Cantor Cutiva and Alex Burdorf. Medical costs and productivity costs related to voice symptoms in colombian teachers. *Journal of Voice*, 29(6):776–e15, 2015.

[17] Harold A Cheyne, Helen M Hanson, Ronald P Genereux, Kenneth N Stevens, and Robert E Hillman. Development and testing of a portable vocal accumulator. *Journal of Speech, Language, and Hearing Research*, 46(6):1457–1467, 2003.

[18] Jan G Švec, Ingo R Titze, and Peter S Popolo. Estimation of sound pressure levels of voiced speech from skin vibration of the neck. *The Journal of the Acoustical Society of America*, 117(3):1386–1394, 2005.

[19] D. D. Mehta, M. Zañartu, S. W. Feng, H. A. Cheyne II, and R. E. Hillman. Mobile voice health monitoring using a wearable accelerometer sensor and a smartphone platform. *IEEE Transactions on Biomedical Engineering*, 59(11):3090–3096, 2012.

[20] Pasquale Bottalico, Arianna Astolfi, and Eric J Hunter. Teachers' voicing and silence periods during continuous speech in classrooms with different reverberation times. *The Journal of the Acoustical Society of America*, 141(1):EL26–EL31, 2017.

[21] Giuseppina Emma Puglisi, Arianna Astolfi, Lady Catherine Cantor Cutiva, and Alessio Carullo. Four-day-follow-up study on the voice monitoring of primary school teachers: Relationships with conversational task and classroom acoustics. *The Journal of the Acoustical Society of America*, 141(1):441–452, 2017.

[22] Annika Szabo Portela, Svante Granqvist, Sten Ternström, and Maria Södersten. Vocal behavior in environmental noise: Comparisons between work and leisure conditions in women with work-related voice disorders and matched controls. *Journal of Voice*, 2017.

[23] Susanna Whitling, Viveka Lyberg-Åhlander, and Roland Rydell. Long-time voice accumulation during work, leisure, and a vocal loading task in groups with different levels of functional voice problems. *Journal of Voice*, 31(2):246–e1, 2017.

[24] Annika Szabo Portela. The female voice in vocally demanding professions: field studies using portable voice accumulators. 2017.

[25] Peter S Popolo, Karen Rogge-Miller, Jan G Svec, and Ingo R Titze. Technical considerations in the design of a wearable voice dosimeter. http://www.ncvs.org/ncvs/library/tech/NCVSOnlineTechnicalMemo05.pdf, 2005. last view: 15/03/2017.

[26] Peter S Popolo, Jan G Svec, and Ingo R Titze. Adaptation of a pocket pc for use as a wearable voice dosimeter. *Journal of Speech, Language, and Hearing Research*, 48(4):780–791, 2005.

[27] Harold A Cheyne, Helen M Hanson, Ronald P Genereux, Kenneth N Stevens, and Robert E Hillman. Development and testing of a portable vocal accumulator. *Journal of Speech, Language, and Hearing Research*, 46(6):1457–1467, 2003.

[28] Robert E Hillman, James T Heaton, Asa Masaki, Steven M Zeitels, and Harold A Cheyne. Ambulatory monitoring of disordered voices. *Annals of Otology, Rhinology & Laryngology*, 115(11):795–801, 2006.

[29] M Wirebrand. Voxlog report. *Stockholm, Sweden, Sweden's Innovation Agency*, 2012.

[30] A. Carullo, A. Vallan, and A. Astolfi. Design issues for a portable vocal analyzer. *IEEE Transactions on Instrumentation and Measurement*, 62(5):1084–1093, 2013.

[31] A Carullo, A Vallan, A Astolfi, L Pavese, and GE Puglisi. Validation of calibration procedures and uncertainty estimation of contact-microphone based vocal analyzers. *Measurement*, 74:130–142, 2015.

[32] Estelle Campione and Jean Véronis. A large-scale multilingual study of silent pause duration. In *Speech prosody 2002, international conference*, 2002.

[33] Parakrant Sarkar and K Sreenivasa Rao. Modeling pauses for synthesis of storytelling style speech using unsupervised word features. *Procedia Computer Science*, 58:42–49, 2015.

[34] Peter S Popolo, Karen Rogge-Miller, Jan G Svec, and Ingo R Titze. Technical considerations in the design of a wearable voice dosimeter. *Journal of the Acoustical Society of America*, 112(5):2304, 2002.

[35] Susanna Simberg, Eeva Sala, Kirsti Vehmas, and Anneli Laine. Changes in the prevalence of vocal symptoms among teachers during a twelve-year period. *Journal of Voice*, 19(1):95–102, 2005.

[36] Lady Catherine Cantor Cutiva, Ineke Vogel, and Alex Burdorf. Voice disorders in teachers and their associations with work-related factors: a systematic review. *Journal of Communication Disorders*, 46(2):143–155, 2013.

[37] Juha Vintturi, Paavo Alku, Eeva Sala, Marketta Sihvo, and Erkki Vilkman. Loading-related subjective symptoms during a vocal loading test with special reference to gender and some ergonomic factors. *Folia phoniatrica et logopaedica*, 55(2):55–69, 2003.

[38] Anne-Maria Laukkanen, Irma Ilomäki, Kirsti Leppänen, and Erkki Vilkman. Acoustic measures and self-reports of vocal fatigue by female teachers. *Journal of Voice*, 22(3):283–289, 2008.

[39] Eric J Hunter and Ingo R Titze. Variations in intensity, fundamental frequency, and voicing for teachers in occupational versus nonoccupational settings. *Journal of Speech, Language, and Hearing Research*, 53(4):862–875, 2010.

[40] Annika Szabo Portela, Britta Hammarberg, and Maria Södersten. Speaking fundamental frequency and phonation time during work and leisure time in vocally healthy preschool teachers measured with a voice accumulator. *Folia Phoniatrica et Logopaedica*, 65(2):84–90, 2013.

[41] Leena Rantala, Erkki Vilkman, and Risto Bloigu. Voice changes during work: subjective complaints and objective measurements for female primary and secondary schoolteachers. *Journal of voice*, 16(3):344–355, 2002.

[42] Eric J Hunter and Ingo R Titze. Variations in intensity, fundamental frequency, and voicing for teachers in occupational versus nonoccupational settings. *Journal of Speech, Language, and Hearing Research*, 53(4):862–875, 2010.

[43] Jan G. Švec, Peter S. Popolo, and Ingo R. Titze. Measurement of vocal doses in speech: Experimental procedure and signal processing. *Logopedics Phoniatrics Vocology*, 28(4):181–192, 2003.

[44] Erkki Vilkman, Eija-Riitta Lauri, Paavo Alku, Eeva Sala, and Marketta Sihvo. Effects of prolonged oral reading on f0, spl, subglottal pressure and amplitude characteristics of glottal flow waveforms. *Journal of Voice*, 13(2):303–312, 1999.

[45] Maria Södersten, Svante Granqvist, Britta Hammarberg, and Annika Szabo. Vocal behavior and vocal loading factors for preschool teachers at work studied with binaural dat recordings. *Journal of Voice*, 16(3):356–371, 2002.

[46] Harlan Lane and Bernard Tranel. The lombard sign and the role of hearing in speech. *Journal of Speech, Language, and Hearing Research*, 14(4):677–709, 1971.

[47] H Lazarus. Prediction of verbal communication is noise—a review: Part 1. *Applied Acoustics*, 19(6):439–464, 1986.

[48] Pasquale Bottalico and Arianna Astolfi. Investigations into vocal doses and parameters pertaining to primary school teachers in classrooms. *The Journal of the Acoustical Society of America*, 131(4):2817–2827, 2012.

[49] Hiroshi Sato and John S Bradley. Evaluation of acoustical conditions for speech communication in working elementary school classrooms. *The Journal of the Acoustical Society of America*, 123(4):2064–2077, 2008.

[50] Bridget Shield, Robert Conetta, Julie Dockrell, Daniel Connolly, Trevor Cox, and Charles Mydlarz. A survey of acoustic conditions and noise levels in secondary school classrooms in england. *The Journal of the Acoustical Society of America*, 137(1):177–188, 2015.

[51] Eeva Sala and Leena Rantala. Acoustics and activity noise in school classrooms in finland. *Applied Acoustics*, 114:252–259, 2016.

[52] Par Lundquist, Kjell Holmberg, and Ulf Landstrom. Annoyance and effects on work from environmental noise at school. *Noise & health*, 2(8):39, 2000.

[53] Yasar Avsar and M Talha Gonullu. The influence of indoor acoustical parameters on student perception in classrooms. *Noise Control Engineering Journal*, 58(3):310–318, 2010.

[54] Arianna Astolfi and Franco Pellerey. Subjective and objective assessment of acoustical and overall environmental quality in secondary school classrooms. *The Journal of the Acoustical Society of America*, 123(1):163–173, 2008.

[55] Jonas Brunskog, Anders Christian Gade, Gaspar Payá Bellester, and Lilian Reig Calbo. Increase in voice level and speaker comfort in lecture rooms. *The Journal of the Acoustical Society of America*, 125(4):2072–2082, 2009.

[56] David Pelegrín-García, Bertrand Smits, Jonas Brunskog, and Cheol-Ho Jeong. Vocal effort with changing talker-to-listener distance in different acoustic environments. *The Journal of the Acoustical Society of America*, 129(4):1981–1990, 2011.

[57] David Pelegrín-García, Jonas Brunskog, Viveka Lyberg-Åhlander, and Anders Löfqvist. Measurement and prediction of voice support and room gain in school classrooms. *The Journal of the Acoustical Society of America*, 131(1):194–204, 2012.

[58] David Pelegrín-García. Comment on "increase in voice level and speaker comfort in lecture rooms"[j. acoust. soc. am. 125, 2072–2082 (2009)](l). *The Journal of the Acoustical Society of America*, 129(3):1161–1164, 2011.

[59] Jan G Švec and Svante Granqvist. Guidelines for selecting microphones for human voice production research. *American Journal of Speech-Language Pathology*, 19(4):356–368, 2010.

[60] Antonella Castellana, Alessio Carullo, Arianna Astolfi, Giuseppina Emma Puglisi, and Umberto Fugiglando. Intra-speaker and inter-speaker variability in speech sound pressure level across repeated readings. *The Journal of the Acoustical Society of America*, 141(4):2353–2363, 2017.

[61] Acoustical Society of America, New York. *ANSI S3.5-1997, Methods for Calculation of the Speech Intelligibility Index*, 1997.

[62] IR Titze. Principles of voice production prentice hall. *Englewood Cliffs, NJ*, 1994.

[63] P Dejonckere. Principal components in voice pathology. *Voice*, 4:96–105, 1995.

[64] American National Standards Institute, New York. *USA Standard, acoustical terminology (S1.1)*, 1960.

[65] David Pisoni and Robert Remez. *The handbook of speech perception*. John Wiley & Sons, 2008.

[66] PH Dejonckere. Perceptual and laboratory assessment of dysphonia. *Otolaryngologic Clinics of North America*, 33(4):731–750, 2000.

[67] Claudia Manfredi and Philippe H Dejonckere. Voice dosimetry and monitoring, with emphasis on professional voice diseases: Critical review and framework for future research. *Logopedics Phoniatrics Vocology*, 41(2):49–65, 2016.

[68] Norman D Hogikyan. The voice-related quality of life (v-rqol) measure: History and ongoing utility of. *SIG 3 Perspectives on Voice and Voice Disorders*, 14:3–5, 2004.

[69] Barbara H Jacobson, Alex Johnson, Cynthia Grywalski, Alice Silbergleit, Gary Jacobson, Michael S Benninger, and Craig W Newman. The voice handicap index (vhi): development and validation. *American Journal of Speech-Language Pathology*, 6(3):66–70, 1997.

[70] Estella PM Ma and Edwin ML Yiu. Voice activity and participation profile: assessing the impact of voice disorders on daily activities. *Journal of Speech, Language, and Hearing Research*, 44(3):511–524, 2001.

[71] Minoru Hirano. Clinical examination of voice. *Disorders of human communication*, 5:1–99, 1981.

[72] Britta Hammarberg. Voice research and clinical needs. *Folia phoniatrica et logopaedica*, 52(1-3):93–102, 2000.

[73] R Hillman. Overview of the consensus auditory-perceptual evaluation of voice (cape-v) instrument developed by asha special interest division 3. In *Annual Symposium on the Care of the Professional Voice, Philadelphia*, 2003.

[74] Bryan Gick, Ian Wilson, and Donald Derrick. *Articulatory Phonetics*. Wiley-Blackwell, 2013.

[75] Ingo R Titze. *Workshop on acoustic voice analysis: Summary statement*. National Center for Voice and Speech, 1995.

[76] S Awan, J Barkmeier-Kraemer, M Courey, D Deliyski, T Eadie, J Svec, and D Paul. Standard clinical protocols for endoscopic, acoustic, and aerodynamic voice assessment: Recommendations from asha expert committee. In *2014 ASHA Convention*, 2014.

[77] Robert E Hillman, Eva B Holmberg, Joseph S Perkell, Michael Walsh, and Charles Vaughan. Objective assessment of vocal hyperfunction: An experimental framework and initial results. *Journal of Speech, Language, and Hearing Research*, 32(2):373–392, 1989.

[78] Eugene H Buder. Acoustic analysis of voice quality: A tabulation of algorithms 1902–1990. *Voice quality measurement*, pages 119–244, 2000.

[79] Steven Bielamowicz, Jody Kreiman, Bruce R Gerratt, Marc S Dauer, and Gerald S Berke. Comparison of voice analysis systems for perturbation measurement. *Journal of Speech, Language, and Hearing Research*, 39(1):126–134, 1996.

[80] Yu Zhang and Jack J Jiang. Acoustic analyses of sustained and running voices from patients with laryngeal pathologies. *Journal of Voice*, 22(1):1–9, 2008.

[81] Soren Y Lowell, Raymond H Colton, Richard T Kelley, and Youngmee C Hahn. Spectral-and cepstral-based measures during continuous speech: capacity to distinguish dysphonia and consistency within a speaker. *Journal of Voice*, 25(5):e223–e232, 2011.

[82] James Hillenbrand and Robert A Houde. Acoustic correlates of breathy vocal quality: dysphonic voices and continuous speech. *Journal of Speech, Language, and Hearing Research*, 39(2):311–321, 1996.

[83] James Hillenbrand, Ronald A Cleveland, and Robert L Erickson. Acoustic correlates of breathy vocal quality. *Journal of Speech, Language, and Hearing Research*, 37(4):769–778, 1994.

[84] Youri Maryn, Nelson Roy, Marc De Bodt, Paul Van Cauwenberge, and Paul Corthals. Acoustic measurement of overall voice quality: A meta-analysis a. *The Journal of the Acoustical Society of America*, 126(5):2619–2634, 2009.

[85] Yolanda D Heman-Ackah, Deirdre D Michael, Margaret M Baroody, Rosemary Ostrowski, James Hillenbrand, Reinhardt J Heuer, Michelle Horman, and Robert T Sataloff. Cepstral peak prominence: a more reliable measure of dysphonia. *Annals of Otology, Rhinology & Laryngology*, 112(4):324–333, 2003.

[86] Cornelia Moers, Bernd Möbius, Frank Rosanowski, Elmar Nöth, Ulrich Eysholdt, and Tino Haderlein. Vowel-and text-based cepstral analysis of chronic hoarseness. *Journal of Voice*, 26(4):416–424, 2012.

[87] Lilia F Brinca, Ana Paula F Batista, Ana Inês Tavares, Ilídio C Gonçalves, and Maria L Moreno. Use of cepstral analyses for differentiating normal from dysphonic voices: A comparative study of connected speech versus sustained vowel in european portuguese female speakers. *Journal of Voice*, 28(3):282–286, 2014.

[88] Youri Maryn, Marc De Bodt, and Nelson Roy. The acoustic voice quality index: toward improved treatment outcomes assessment in voice disorders. *Journal of communication disorders*, 43(3):161–174, 2010.

[89] P. Boersma and D. Weenink. Praat. http://www.praat.org/. last view: 15/03/2017.

[90] J.M. Hillenbrand. Speech tool. http://homepages.wmich.edu/ hillenbr/. last view: 15/03/2017.

[91] Shaheen N Awan, Nelson Roy, Marie E Jetté, Geoffrey S Meltzner, and Robert E Hillman. Quantifying dysphonia severity using a spectral/cepstral-based acoustic index: comparisons with auditory-perceptual judgements from the cape-v. *Clinical linguistics & phonetics*, 24(9):742–758, 2010.

[92] Daryush D Mehta, Jarrad H Van Stan, and Robert E Hillman. Relationships between vocal function measures derived from an acoustic microphone and a subglottal neck-surface accelerometer. *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, 24(4):659–668, 2016.

[93] Claudia Manfredi, Tommaso Bruschi, Alessandro Dallai, Alessandro Ferri, Piero Tortoli, and Marcello Calisti. Voice quality monitoring: a portable device prototype. In *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, pages 997–1000. IEEE, 2008.

[94] C Manfredi, J Lebacq, G Cantarella, J Schoentgen, S Orlandi, A Bandini, and PH DeJonckere. Smartphones offer new opportunities in clinical voice research. *Journal of Voice*, 31(1):111–e1, 2017.

[95] Daryush D Mehta, Matías Zañartu, Jarrad H Van Stan, Shengran W Feng, Harold A Cheyne, and Robert E Hillman. Smartphone-based detection of voice disorders by long-term monitoring of neck acceleration features. In

Body Sensor Networks (BSN), 2013 IEEE International Conference on, pages 1–6. IEEE, 2013.

[96] Eva Van Leer, Robert C Pfister, and Xuefu Zhou. An ios-based cepstral peak prominence application: feasibility for patient practice of resonant voice. *Journal of Voice*, 31(1):131–e9, 2017.

[97] Jarrad H. Van Stan, Daryush D. Mehta, Steven M. Zeitels, James A. Burns, Anca M. Barbu, and Robert E. Hillman. Average ambulatory measures of sound pressure level, fundamental frequency, and vocal dose do not differ between adult females with phonotraumatic lesions and matched control subjects. *Annals of Otology, Rhinology & Laryngology*, 124(11):864–874, 2015.

[98] Marzyeh Ghassemi, Jarrad H Van Stan, Daryush D Mehta, Matías Zañartu, Harold A Cheyne, Robert E Hillman, and John V Guttag. Learning to detect vocal hyperfunction from ambulatory neck-surface acceleration features: Initial results for vocal fold nodules. *IEEE Transactions on Biomedical Engineering*, 61(6):1668–1675, 2014.

[99] Arianna Astolfi, Alessio Carullo, Simone Corbellini, Massimo Spadola, Anna Accornero, Giuseppina E Puglisi, Antonella Castellana, Louena Shtrepi, Gian Luca D'Antonio, Alessandro Peretti, et al. Long-term voice monitoring with smartphone applications and contact microphone. *The Journal of the Acoustical Society of America*, 141(5):3541–3541, 2017.

[100] World Health Organization. *International Classification of Functioning, Disability and Health: ICF*. World Health Organization, 2001.

[101] Harold A Cheyne, Kaustubh Kalgaonkar, Mark Clements, and Patrick Zurek. Talker-to-listener distance effects on speech production and perception. *The Journal of the Acoustical Society of America*, 126(4):2052–2060, 2009.

[102] David Pelegrín-García, Bertrand Smits, Jonas Brunskog, and Cheol-Ho Jeong. Vocal effort with changing talker-to-listener distance in different acoustic environments. *The Journal of the Acoustical Society of America*, 129(4):1981–1990, 2011.

[103] John W Black. The effect of room characteristics upon vocal intensity and rate. *The Journal of the Acoustical Society of America*, 22(2):174–176, 1950.

[104] Alessio Carullo, Federico Casassa, Antonella Castellana, Arianna Astolfi, Lorenzo Pavese, and Giuseppina Emma Puglisi. Performance comparison of different contact microphones used for voice monitoring. In *22nd Interntional Congress on Sound and Vibration*. IIAV, 2015.

[105] International Organization for Standardization, Geneva, Switzerland. *BS EN ISO 3382-2, Acoustics—Measurement of Room Acoustic Parameters—Part 2: Reverberation Time in Ordinary Rooms*, 2008.

[106] Deutsche Institutfür Normung, Berlin. *DIN 18041: Hörsamkeit in Räumen—Anforderungen, Empfehlungen und Hinweisefür die Planung (Acoustic Quality in Rooms- Specifications and Instructions for the Room Acoustic Design)*, 2016.

[107] International Organization for Standardization, Geneva, Switzerland. *ISO 1996, Acoustics—Description, Measurements and Assessment of Environmental Noise*, 2016.

[108] Giuseppina Emma Puglisi, Arianna Astolfi, Lady Catherine Cantor Cutiva, and Alessio Carullo. Assessment of indoor ambient noise level in school classrooms. In *EuroNoise 2015*, pages 715–720. EAA-NAG-ABAV, 2015.

[109] Nick Durup, Bridget Shield, Stephen Dance, and Rory Sullivan. An investigation into relationships between classroom acoustic measurements and voice parameters of teachers. *Building Acoustics*, 22(3-4):225–241, 2015.

[110] Brian S Everitt and David C Howell. *Encyclopedia of statistics in behavioral science.* John Wiley & Sons Ltd, 2005.

[111] Pasquale Bottalico, Ivano Ipsaro Passione, Simone Graetzer, and Eric J Hunter. Evaluation of the starting point of the lombard effect. *Acta Acustica united with Acustica*, 103(1):169–172, 2017.

[112] Lau Nijs and Monika Rychtáriková. Calculating the optimum reverberation time and absorption coefficient for good speech intelligibility in classroom design using u50. *Acta Acustica united with Acustica*, 97(1):93–102, 2011.

[113] W Yang and JS Bradley. Effects of room acoustics on the intelligibility of speech in classrooms for young children. *The Journal of the Acoustical Society of America*, 125(2):922–933, 2009.

[114] Department for Education and the Educational Funding Agency, London. *WSP Bulletin BB93, Acoustic Design in Schools: Performance Standards Building Bulletin*, 2015.

[115] French Ministry of Ecology and Sustainable Development, Paris, France. *Circular of 25 April 2003 Relative to the Application of Acoustic Regulation for Buildings Other Than Apartment Buildings*, 2003.

[116] EU GPP, Brussels, Belgium. *Acustica in Edilizia. Classificazione Acustica delle Unità Immobiliari. Procedura di Valutazione e Verifica in Opera (Building Acoustics. Acoustic Classification of Building Units. Evaluation Procedure and In Situ Measurements)*, 2016.

[117] E.B. Holmberg, P. Doyle, J.S. Perkell, B. Hammarberg, and R.E. Hillman. Aerodynamic and acoustic voice measurements of patients with vocal nodules: variation in baseline and changes across voice therapy. *Journal of Voice*, 17(3):269–282, 2003.

[118] M. Hirano, S. Tanaka, M. Fujita, and R. Terasawa. Fundamental frequency and sound pressure level of phonation in pathological states. *Journal of Voice*, 5(2):120–127, 1991.

[119] Eva B. Holmberg, Robert E. Hillman, Britta Hammarberg, Maria Södersten, and Patricia Doyle. Efficacy of a behaviorally based voice therapy protocol for vocal nodules. *Journal of Voice*, 15(3):395 – 412, 2001.

[120] Takashi Masuda, Yoshimitu Ikeda, Hiroko Manako, and Sohtaro Komiyama. Analysis of vocal abuse: Fluctuations in phonation time and intensity in 4 groups of speakers. *Acta Oto-Laryngologica*, 113(4):547–552, 1993.

[121] Elizabeth Erickson Levendoski, Ciara Leydon, and Susan L. Thibeault. Vocal fold epithelial barrier in health and injury a research review. *J Speech Lang Hear Res.*, 57(5):1679–1691, 2014.

[122] Hartmut Traunmüller and Anders Eriksson. Acoustic effects of variation in vocal effort by men, women, and children. *The Journal of the Acoustical Society of America*, 107(6):3438–3451, 2000.

[123] Nick Durup, Bridget Shield, Stephen Dance, and Rory Sullivan. An investigation into relationships between classroom acoustic measurements and voice parameters of teachers. *Building Acoustics*, 22(3-4):225–241, 2015.

[124] Murray Hodgson, Rod Rempel, and Susan Kennedy. Measurement and prediction of typical speech and background-noise levels in university classrooms during lectures. *The Journal of the Acoustical Society of America*, 105(1):226–233, 1999.

[125] Lady Catherine Cantor Cutiva, Giuseppina Emma Puglisi, Arianna Astolfi, and Alessio Carullo. Four-day follow-up study on the self-reported voice condition and noise condition of teachers: relationship between vocal parameters and classroom acoustics. *Journal of Voice*, 31(1):120–e1, 2017.

[126] Arianna Astolfi, Alessio Carullo, Lorenzo Pavese, and Giuseppina Emma Puglisi. Duration of voicing and silence periods of continuous speech in different acoustic environments. *The Journal of the Acoustical Society of America*, 137(2):565–579, 2015.

[127] Martin Cooke, Simon King, Maëva Garnier, and Vincent Aubanel. The listening talker: A review of human and algorithmic context-induced modifications of speech. *Computer Speech & Language*, 28(2):543–571, 2014.

[128] Harlan Lane, Bernard Tranel, and Cyrus Sisson. Regulation of voice communication by sensory dynamics. *The Journal of the Acoustical Society of America*, 47(2B):618–624, 1970.

[129] Ian R Cushing, Francis F Li, Trevor J Cox, Ken Worrall, and Tim Jackson. Vocal effort levels in anechoic conditions. *Applied Acoustics*, 72(9):695–701, 2011.

[130] Wayne O Olsen. Average speech levels and spectra in various speaking/listening conditions: A summary of the pearson, bennett, & fidell (1977) report. *American Journal of Audiology*, 7(2):21–25, 1998.

[131] Michael A Picheny, Nathaniel I Durlach, and Louis D Braida. Speaking clearly for the hard of hearing ii: Acoustic characteristics of clear and conversational speech. *Journal of speech and hearing research*, 29(4):434–446, 1986.

[132] Denis Byrne, Harvey Dillon, Khanh Tran, Stig Arlinger, Keith Wilbraham, Robyn Cox, Bjorn Hagerman, Raymond Hetu, Joseph Kei, C Lui, et al. An international comparison of long-term average speech spectra. *The Journal of the Acoustical Society of America*, 96(4):2108–2120, 1994.

[133] Eva B Holmberg, Robert E Hillman, Joseph S Perkell, and Carla Gress. Relationships between intra-speaker variation in aerodynamic measures of voice production and variation in spl across repeated recordings. *Journal of Speech, Language, and Hearing Research*, 37(3):484–495, 1994.

[134] WS Brown Jr, Thomas Murry, and David Hughes. Comfortable effort level: an experimental variable. *The Journal of the Acoustical Society of America*, 60(3):696–699, 1976.

[135] WS Brown, Richard J Morris, and Thomas Murry. Comfortable effort level revisited. *Journal of Voice*, 10(3):299–305, 1996.

[136] Kathryn L Garrett and E Charles Healey. An acoustic analysis of fluctuations in the voices of normal adult speakers across three times of day. *The Journal of the Acoustical Society of America*, 82(1):58–62, 1987.

[137] Paul Corthals. Sound pressure level of running speech: percentile level statistics and equivalent continuous sound level. *Folia phoniatrica et logopaedica*, 56(3):170–181, 2004.

[138] Marketta Sihvo and Eeva Sala. Sound level variation findings for pianissimo and fortissimo phonations in repeated measurements. *Journal of Voice*, 10(3):262–268, 1996.

[139] Marketta Sihvo, Pekka Laippala, and Eeva Sala. A study of repeated measures of softest and loudest phonations. *Journal of Voice*, 14(2):161–169, 2000.

[140] Eva B Holmberg, Joseph S Perkell, Robert E Hillman, and Carla Gress. Individual variation in measures of voice. *Phonetica*, 51(1-3):30–37, 1994.

[141] Ingo R Titze. On the relation between subglottal pressure and fundamental frequency in phonation. *The Journal of the Acoustical Society of America*, 85(2):901–906, 1989.

[142] Joint Committee for Guides in Metrology. Jcgm 100, evaluation of measurement data - guide to the expression of uncertainty in measurement. http://www.bipm.org/en/publications/guides, 2008. last view: 15/03/2017.

[143] Jacek Szudek, Amberley Ostevik, Peter Dziegielewski, Jason Robinson-Anagor, Nahla Gomaa, Bill Hodgetts, Allan Ho, C Mathers, A Smith, M Concha, et al. Can uhear me now? validation of an ipod-based hearing loss screening test. *Journal of Otolaryngology-Head and Neck Surgery*, 41(1):S78, 2012.

[144] James C Wang, Steven Zupancic, Coby Ray, Joehassin Cordero, and Joshua C Demke. Hearing test app useful for initial screening, original research shows. *The Hearing Journal*, 67(10):32–34, 2014.

[145] I. Vernero, M. Gambino, R. Stefanin, and O. Schindler. *Cartella Logopedica. Età Evolutiva (Logopedic Collection. Childhood)*. Edizioni Omega, Torino, 1998.

[146] H.K. Schutte and W. Seidner. Recommendation by the union of european phoniatricians (uep): Standardizing voice area measurement/phonetography. *Folia Phoniatrica et Logopaedica*, 35:286–288, 1983.

[147] Hana Šrámková, Svante Granqvist, Christian T Herbst, and Jan G Švec. The softest sound levels of the human voice in normal subjects. *The Journal of the Acoustical Society of America*, 137(1):407–418, 2015.

[148] Wouter A Dreschler, Hans Verschuure, Carl Ludvigsen, and Søren Westermann. Icra noises: Artificial noise signals with speech-like spectral and temporal properties for hearing instrument assessment: Ruidos icra: Señates de ruido artificial con espectro similar al habla y propiedades temporales para pruebas de instrumentos auditivos. *Audiology*, 40(3):148–157, 2001.

[149] J.D. Gibbons and S. Chakraborti. *Nonparametric Statistical Inference*. Taylor 'I&' Francis, 2003.

[150] Federico Casassa, Alessandro Schiavi, and Adriano Troia. Development of a test system for voice monitoring contact sensor: phonatory system simulator. In *24nd International Congress on Sound and Vibration*. IIAV, 2017.

[151] Martin Cooke, Catherine Mayo, and Julián Villegas. The contribution of durational and spectral changes to the lombard speech intelligibility benefit. *The Journal of the Acoustical Society of America*, 135(2):874–883, 2014.

[152] Jean-Claude Junqua. The lombard reflex and its role on human listeners and automatic speech recognizers. *The Journal of the Acoustical Society of America*, 93(1):510–524, 1993.

[153] Jean-Sylvain Liénard and Maria-Gabriella Di Benedetto. Effect of vocal effort on spectral properties of vowels. *The Journal of the Acoustical Society of America*, 106(1):411–422, 1999.

[154] Deirdre D Michael, Gerald M Siegel, and Herbert L Pick. Effects of distance on vocal intensity. *Journal of Speech, Language, and Hearing Research*, 38(5):1176–1183, 1995.

[155] Viveka Lyberg Åhlander, David Pelegrín García, Susanna Whitling, Roland Rydell, and Anders Löfqvist. Teachers' voice use in teaching environments: a field study using ambulatory phonation monitor. *Journal of Voice*, 28(6):841–e5, 2014.

[156] David Pelegrín-García and Jonas Brunskog. Speakers' comfort and voice level variation in classrooms: Laboratory research. *The Journal of the Acoustical Society of America*, 132(1):249–260, 2012.

[157] Jean C Krause and Louis D Braida. Acoustic properties of naturally produced clear speech at normal speaking rates. *The Journal of the Acoustical Society of America*, 115(1):362–378, 2004.

[158] Valerie Hazan and Rachel Baker. Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions a. *The Journal of the Acoustical Society of America*, 130(4):2139–2152, 2011.

[159] Anne H Anderson, Miles Bader, Ellen Gurman Bard, Elizabeth Boyle, Gwyneth Doherty, Simon Garrod, Stephen Isard, Jacqueline Kowtko, Jan McAllister, Jim Miller, et al. The hcrc map task corpus. *Language and speech*, 34(4):351–366, 1991.

[160] Mike Barron and L-J Lee. Energy relations in concert auditoriums. i. *The Journal of the Acoustical Society of America*, 84(2):618–628, 1988.

[161] Steven B Leder and Jaclyn B Spitzer. Speaking fundamental frequency, intensity, and rate of adventitiously profoundly hearing-impaired adult women. *The Journal of the Acoustical Society of America*, 93(4):2146–2151, 1993.

[162] Maëva Garnier, Nathalie Henrich, and Daniele Dubois. Influence of sound immersion and communicative interaction on the lombard effect. *Journal of Speech, Language, and Hearing Research*, 53(3):588–608, 2010.

[163] Dennis H Klatt. Linguistic uses of segmental duration in english: Acoustic and perceptual evidence. *The Journal of the Acoustical Society of America*, 59(5):1208–1221, 1976.

[164] Carl E Williams and Kenneth N Stevens. Emotions and speech: Some acoustical correlates. *The Journal of the Acoustical Society of America*, 52(4B):1238–1250, 1972.

[165] Jean C Krause and Louis D Braida. Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility. *The Journal of the Acoustical Society of America*, 112(5):2165–2172, 2002.

[166] W Van Summers, David B Pisoni, Robert H Bernacki, Robert I Pedlow, and Michael A Stokes. Effects of noise on speech production: Acoustic and perceptual analyses. *The Journal of the Acoustical Society of America*, 84(3):917–928, 1988.

[167] Marcella Cipriano, Arianna Astolfi, and David Pelegrín-García. Combined effect of noise and room acoustics on vocal effort in simulated classrooms. *The Journal of the Acoustical Society of America*, 141(1):EL51–EL56, 2017.

[168] Pasquale Bottalico, Simone Graetzer, and Eric J Hunter. Effects of speech style, room acoustics, and vocal fatigue on vocal effort. *The Journal of the Acoustical Society of America*, 139(5):2870–2879, 2016.

[169] International Organization for Standardization, Geneva, Switzerland. *ISO 3741, Acoustics—Determination of sound power levels and sound energy levels of noise sources using sound pressure—Precision methods for reverberation test rooms*, 2010.

[170] Guus de Krom. A cepstrum-based technique for determining a harmonics-to-noise ratio in speech signals. *Journal of Speech, Language, and Hearing Research*, 36(2):254–266, 1993.

[171] Huiwen Goy, David N. Fernandes, M. Kathleen Pichora-Fuller, and Pascal van Lieshout. Normative voice data for younger and older adults. *Journal of Voice*, 27(5):545 – 555, 2013.

[172] J.M Hillenbran. cpps.exe. https://homepages.wmich.edu/ hillenbr/. last view: 03/09/2017.

[173] Christopher R Watts, Shaheen N Awan, and Youri Maryn. A comparison of cepstral peak prominence measures from two acoustic analysis programs. *Journal of Voice*, 31(3):387–e1, 2017.

[174] Robert F Coleman. Sources of variation in phonetograms. *Journal of Voice*, 7(1):1–14, 1993.

[175] D. Hosmer, S. Lemeshow, and R. Sturdivant. *Applied Logistic Regression*. Wiley, third edition edition, 2013.

[176] Tolvan Data. Sopran. http://www.tolvan.com/index.php?page=/sopran/sopran.php. last view: 17/06/2017.

[177] Antonella Castellana, Alessio Carullo, Simone Corbellini, Arianna Astolfi, M Spadola Bisetti, and J Colombini. Cepstral peak prominence smoothed distribution as discriminator of vocal health in sustained vowel. In *Instrumentation and Measurement Technology Conference (I2MTC), 2017 IEEE International*, pages 552–557. IEEE, 2017.

[178] Peter J Murphy. On first rahmonic amplitude in the analysis of synthesized aperiodic voice signals. *The Journal of the Acoustical Society of America*, 120(5):2896–2907, 2006.

[179] Rubén Fraile and Juan Ignacio Godino-Llorente. Cepstral peak prominence: A comprehensive analysis. *Biomedical Signal Processing and Control*, 14:42–54, 2014.

[180] Michael HL Hecker and E James Kreul. Descriptions of the speech of patients with cancer of the vocal folds. part i: Measures of fundamental frequency. *The Journal of the Acoustical Society of America*, 49(4B):1275–1282, 1971.

[181] Anders G Askenfelt and Britta Hammarberg. Speech waveform perturbation analysis: A perceptual-acoustical comparison of seven measures. *Journal of Speech, Language, and Hearing Research*, 29(1):50–64, 1986.

[182] Christopher R Watts and Shaheen N Awan. Use of spectral/cepstral analyses for differentiating normal from hypofunctional voices in sustained vowel and continuous speech contexts. *Journal of Speech, Language, and Hearing Research*, 54(6):1525–1537, 2011.

[183] Youri Maryn, Paul Corthals, Paul Van Cauwenberge, Nelson Roy, and Marc De Bodt. Toward improved ecological validity in the acoustic measurement of overall voice quality: combining continuous speech and sustained vowels. *Journal of voice*, 24(5):540–555, 2010.

[184] Youri Maryn, Marc De Bodt, Ben Barsties, and Nelson Roy. The value of the acoustic voice quality index as a measure of dysphonia severity in subjects speaking different languages. *European Archives of Oto-Rhino-Laryngology*, 271(6):1609–1619, 2014.

[185] Cara Sauder, Michelle Bretl, and Tanya Eadie. Predicting voice disorder status from smoothed measures of cepstral peak prominence using praat and analysis of dysphonia in speech and voice (adsv). *Journal of Voice*, 2017.

[186] Youri Maryn and David Weenink. Objective dysphonia measures in the program praat: smoothed cepstral peak prominence and acoustic voice quality index. *Journal of Voice*, 29(1):35–43, 2015.

[187] Gaetano Fava, Nico Paolo Paolillo, Gisele Oliveira, and Mara Behlau. Cross-cultural adaptation, validation, and cutoff point of the italian version of the voice activity and participation profile: Profilo di attività e partecipazione vocale. *Journal of Voice*, 29(1):130–e11, 2015.

[188] Shaheen N Awan, Nelson Roy, and Seth M Cohen. Exploring the relationship between spectral and cepstral measures of voice and the voice handicap index (vhi). *Journal of Voice*, 28(4):430–439, 2014.

[189] Stephen Jannetts and Anja Lowit. Cepstral analysis of hypokinetic and ataxic voices: correlations with perceptual and other acoustic measures. *Journal of Voice*, 28(6):673–680, 2014.

[190] Yolanda D Heman-Ackah, Deirdre D Michael, and George S Goding. The relationship between cepstral peak prominence and selected parameters of dysphonia. *Journal of Voice*, 16(1):20–27, 2002.

[191] Vijay Parsa and Donald G Jamieson. Acoustic discrimination of pathological voice: sustained vowels versus continuous speech. *Journal of Speech, Language, and Hearing Research*, 44(2):327–339, 2001.

[192] Italian speechdat (ii) fdb-3000l. $http://catalog.elra.info/product_info.php?products_id = 631$. last view: 15/03/2017.

[193] Børge Lindberg, Finn Tore Johansen, Narada D Warakagoda, Gunnar Lehtinen, Zdravko Kacic, Andrej Zgank, Kjell Elenius, and Giampiero Salvi. A noise robust multilingual reference recogniser based on speechdat (ii). In *INTERSPEECH*, pages 370–373, 2000.

[194] Joshua S Richman and J Randall Moorman. Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology-Heart and Circulatory Physiology*, 278(6):H2039–H2049, 2000.

[195] Steven M Pincus. Approximate entropy as a measure of system complexity. *Proceedings of the National Academy of Sciences*, 88(6):2297–2301, 1991.

[196] Kathiresan Manickam, Christopher Moore, Terry Willard, and Nicholas Slevin. Quantifying aberrant phonation using approximate entropy in electrolaryngography. *Speech communication*, 47(3):312–321, 2005.

[197] Chiara Fabris, Wladimiro De Colle, and Giovanni Sparacino. Voice disorders assessed by (cross-) sample entropy of electroglottogram and microphone signals. *Biomedical Signal Processing and Control*, 8(6):920–926, 2013.

[198] D.K. Lake, J.R. Moorman, and C. Hanqing. Sampen for matlab 1.1-1. http://www.physionet.org/physiotools/sampen/matlab/1.1-1/. last view: 15/03/2017.

[199] Alain De Cheveigné and Hideki Kawahara. Yin, a fundamental frequency estimator for speech and music. *The Journal of the Acoustical Society of America*, 111(4):1917–1930, 2002.

[200] Mateo Aboy, David Cuesta-Frau, Daniel Austin, and Pau Mico-Tormos. Characterization of sample entropy in the context of biomedical signal analysis. In *Engineering in medicine and biology society, 2007. EMBS 2007. 29th Annual international conference of the IEEE*, pages 5942–5945. IEEE, 2007.

[201] Ingo R Titze. *Workshop on acoustic voice analysis: Summary statement*. National Center for Voice and Speech, 1995.

[202] Viv Bewick, Liz Cheek, and Jonathan Ball. Statistics review 13: receiver operating characteristic curves. *Critical care*, 8(6):508, 2004.

Appendix A

Additional material

A.1 Italian passage P1

Avevo un bulldog che si chiamava Bulka. Era tutto nero salvo una macchia bianca all'estremità delle zampe anteriori. Nei cani di questa razza, la mandibola è sempre prominente, così i denti superiori vengono a collocarsi dietro a quelli inferiori. Ma quella di Bulka era tanto grossa che tra gli uni e gli altri denti rimaneva molto spazio. Aveva il muso largo, grandi occhi neri e brillanti e i canini sempre scoperti, perfettamente bianchi. Somigliava a un grugno. Bulka era assai forte. E se afferrava qualcosa tra i denti non c'era verso che mollasse la sua preda. Stretti i canini nella carne dell'avversario, serrava la mascella e rimaneva sospeso come un cencio ad un chiodo: attaccato come una sanguisuga. Un giorno che era stato lanciato contro un orso, gli afferrò tra i denti un orecchio. L'orso cercava di colpirlo con una zampa, scuoteva la testa, ma non se ne poteva sbarazzare: finì per rovesciare il testone in terra per schiacciarvi il cane. Su quest'ultimo, però, perché lasciasse la presa, dovemmo gettare una secchia di acqua gelata. Lo avevo avuto da ragazzo e gli davo da mangiare io stesso. Quando dovetti partire a prestar servizio ne Caucaso, decisi di non prenderlo con me e cercai di andarmene senza che lo sapesse. Ordinai che lo tenessero rinchiuso. Ero giunto alla prima tappa, stavo per ripartire con i cavalli freschi, quando ad un tratto notai una palla nera e brillante che avanzava velocissima sulla strada. Era Bulka col suo collare di rame al collo. Correva a perdifiato; si gettò su di me, mi leccò la mano e poi, la lingua ciondoloni, si stese all'ombra sotto

la vettura. Seppi più tardi che aveva rotto un vetro per seguirmi; era saltato dalla finestra: aveva percorso venti chilometri d'estate, sotto un sole bruciante.

A.2 Italian passage P2

Il papà (o il babbo come dice il piccolo Dado) era sul letto. Sotto di lui, accanto al lago, sedeva Gigi detto Ciccio, cocco della mamma e della nonna. Vicino ad un sasso c'era una rosa rosso vivo e, lo sciocco, vedendola, la volle per la zia. La zia Lulù cercava zanzare per il suo ramarro, ma dato che era giugno (o luglio non so bene) non ne trovava. Trovò invece una rana, che, saltando dalla strada, finì nel lago con un grande spruzzo. Sai che fifa, la zia! Lo schizzo bagnò il suo completo rosa che divenne giallo come un taxì. Passava di lì un signore cosmopolita di nome Sardanapalo Nabucodonosor che si innamorò della zia e la portò con sé in Afghanistan.

A.3 Swedish passage

En pojke kom en dag inspringande på en bondgård och undrade, om han kunde få låna en spade. När bonden frågade, vad han skulle ha den till, svarade pojken att hans bror hade ramlat i ett träsk och att han måste gräva upp honom. - Hur djupt har han ramlat i? frågade bonden. - Upp till vristerna, blev svaret. - Men då han han väl gå därifrån utan din hjälp. Då behöver du väl ingen spade? Pojken såg förtvivlad ut och sa: - Jo, men ni förstår, han ramlade i med huvet först